UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Physics for life sciences**

Arrêté ministérial : 7 août 2006

Présentée par

Thomas Cutuil

Thèse dirigée par **Bernhard Brutscher**

préparée au sein du **Biomolecular NMR group, Institut de Biologie Structurale**
et de l' **Ecole Doctorale de Physique**

Folding of proteins studied by real time NMR and other biophysical methods: the example of β_2-microglobulin

Thèse soutenue publiquement le **14 Décembre 2012**,
devant le jury composé de :

Dr. Sophie Zinn-Justin
DR-CEA (CEA Saclay, France), Rapporteur
Pr. Jochen Balbach
Professeur (Université de Halle, Allemagne), Rapporteur
Pr. Alessandra Corazza
Professeur (Université d'Udine, Italie), Examinatrice
Dr. Vincent Forge
DR-CEA (IRTSV, CEA Grenoble, France), Examinateur
Dr. Dominique Housset
DR-CEA (IBS Grenoble, France), Examinateur
Dr. Bernhard Brutscher
DR-CEA (IBS Grenoble, France), Directeur de thèse

Contents

Nomenclature

AFM	Atomic force microscopy
ANS	8-Anilinonaphthalene-1-sulfonic acid, a fluorescent molecular probe that binds to hydrophobic patches of proteins
ATR	Attenuated total reflectance
B2M	Beta-2-microglobulin
CD	Circular Dichroism
DNA	Desoxyribonucleic Acid
DRA	Dialysis Related Amyloidosis
DTT	Dithiothréitol
ESI-MS	Electrospray ionization mass spectrometry
FT-IR	Fourier Transform Infrared Spectroscopy
HLA	Human Leukocyte Antigens
HMQC	Heteronuclear multiple quantum coherence
HSQC	Heteronuclear single quantum coherence
I*	NMR-invisible folding intermediate state of B2M
I_1	Monomeric folding intermediate state of B2M
I_2	Low molecular weight oligomeric folding intermediate state of B2M, likely to be a dimer
I_3	Short-lived, heavy molecular weight oligomeric folding intermediate state of B2M
IAPP	Islet amyloid polypeptide or amylin

I_t	Folding intermediate state of B2M, with a trans peptide bond before Proline 32
M*	Activated intermediate state on the oligomerization pathway of B2M
MHC	Major Histocompatibility Complex
NMR	Nuclear Magnetic Resonance
PBS	Phosphate Buffer Saline
SAA	Serum Amyloid A
SAXS	Small Angle Xray Scattering
SDS	Sodium Dodecyl Sulfate
BEST	Band-selective Excitation Short-Transient
SOFAST	band-Selective Optimized Flip-Angle Short-Transient
τ_c	Molecular tumbling correlation time
TEM	Transmission electron microscopy
ThT	Thioflavin-T
TROSY	Transverse relaxation optimized spectroscopy
UV	Ultra-Violet
WT	Wild type

Part I

The protein folding problem

Chapter 1

General principles

Protein folding is the process through which a protein reaches its stable, functional, conformation. In his Nobel Prize winning work, Anfinsen states that this stable conformation is unique, and determined only by the amino-acid sequence of the protein, at least for small proteins. This links the genetic coding of a function as a sequence in DNA to its material mode of action which is the lock-key interaction, through a unique 3D protein structure. This of course should be moderated, as the structure of proteins, even in their native functional state, should be considered as an ensemble of highly similar states, rather than as a unique, static structure. Moreover, more and more intrinsically disordered proteins are being discovered, which are proteins that are functional despite their lack of well defined structure. Some of them will even acquire a defined tertiary structure upon binding with another protein or substrate.

The traditional view of protein folding, the acquisition of this stable, functional structure, derives from the Levinthal paradox, which states that if a polypeptide chain should sequentially experience every conformation possible before reaching the functional tertiary structure, implies a folding time that is longer than the age of the universe. Yet, most small proteins are folding within milliseconds. This paradox suggests that proteins do not sample all possible conformations when folding, but rather follow a preferential folding pathway. This can be illustrated using the folding funnel model, meaning that the energy landscape of accessible protein conformations is funnel-shaped, with the native folded structure being a deep, steep-walled minimum. The funnel energy landscape model doesn't mean that the folding is straightforward, limited to a two-state transition between the unfolded and folded structures. The protein can be kinetically trapped in many local energy minima along the folding pathway, and sometimes needs the intervention of chaperone proteins to go on with its folding. An illustration of such a rough folding funnel is shown on figure 1.1.

Protein folding is thought to be mainly driven by the hydrophobic effect, corresponding to the burying of hydrophobic lateral chain amino-acids inside a hydrophobic

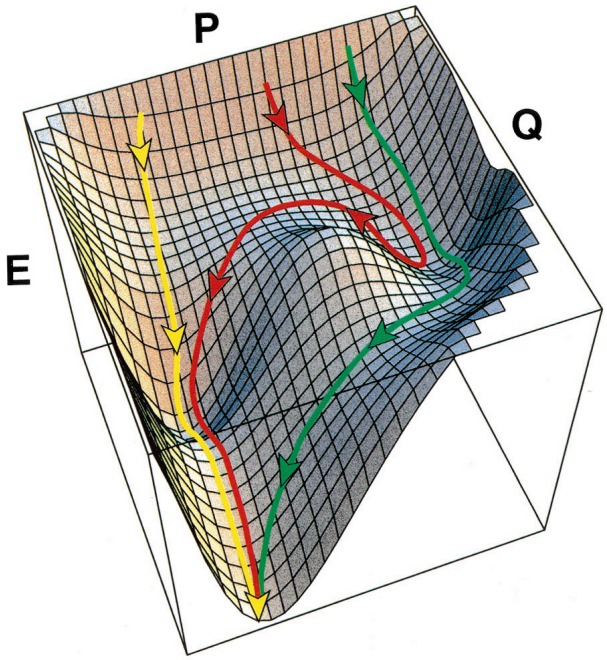


Figure 1.1: Model of the folding landscape of lyzozyme. Unfolded conformations are shown with a higher free energy, on the z axis. The native conformation is at the bottom of the funnel, while two local minima can be seen, corresponding to two folding intermediate states. Three folding pathways are shown, sampling either or both intermediate conformations. Reproduced from [48]

core of the protein. The hydrophobic effect is not due to the interaction between these hydrophobic lateral chains, but is entropy-driven. When exposed to water, in the unfolded conformation, a hydrophobic lateral chain disrupts the highly dynamic hydrogen bond network between water molecules. This hydrogen bond network is restored through the formation of a hydration shell around the hydrophobic chain, but water molecules in this shell have a reduced mobility, equivalent to strongly reduced entropy. The burying of the hydrophobic lateral chain inside the core of the protein releases these water molecules, leading to an increase in entropy of the whole system, and thus a reduced free energy. Of course other interactions are involved in folding, such as hydrogen bonds between amino-acids, salt bridges or hydrophobic, especially aromatic, interactions and disulfide bridges.

The major limitation to the study of protein folding is its timescale: micro- to milliseconds for most small (∼5-50kDa) proteins. Traditional physical methods that are used to study the structure of proteins have time resolutions that are at best of the order of seconds, if at all adapted to the detection of transient species. One solution to this problem is the modification of the folding conditions, at the expense of physiological relevance: a lower temperature slows down the folding often several orders of magnitude, a change of the solvent can also have an effect on the folding speed.

Another solution is reducing the experimental time needed to acquire signals with the different methods, and the development of stopped-flow systems in UV spectroscopy and circular dichroism and high speed pulse sequences in NMR is a way of tackling this difficulty. Finally biochemical modifications of the protein, such as mutation of one or several amino-acids, can trap it in non native conformations, producing equilibrium models of transient species. Yet most studies focusing on protein folding are using proteins that naturally fold at much lower timescales, of the order of minutes. Luckily, some of them are also linked to misfolding diseases.

Misfolded proteins are usually detected by the cell and degraded through proteolysis, but some diseases are due to mutated proteins that cannot reach their native folded state, or to the adoption by wild-type proteins of non-native conformations in specific conditions. Disfunctioning in the proteolytic system can also lead to the accumulation of misfolded protein, which can be pathogenic. Misfolding diseases can be either due to the loss of function of a misfolded protein, or to the accumulation of misfolded protein aggregates [41]. For example, the misfolding of a tumor suppressing protein, such as p53, can cause cancers, while the accumulation of beta and tau aggregates, in the form of amyloids, is characteristic of Alzheimer's disease. This involvement of misfolding in many diseases gives to the study of protein folding a direct medical relevance.

Chapter 2

Amyloidoses as misfolding diseases

2.1 Amyloidosis

Amyloidosis is defined as a group of medical conditions in which normally innocuous soluble proteins polymerize to form insoluble fibrils [156], making of amyloidosis a class of misfolding diseases. Amyloid fibrils grow and associate with constituents of extracellular medium such as plasma, matrix proteins, proteoglycans, to form amyloid deposits in tissues, destroying their normal architecture and function. Amyloid deposits are involved in a lot of very different medical conditions such as some cancers, Alzheimer's disease, rheumatoid arthritis, and Parkinson's disease [69]. Amyloidoses can be systemic, like the Dialysis Related Amyloidosis (DRA), or localized in a single organ, such as brain for Alzheimer's disease. Systemic amyloidoses are based on plasma circulating precursors, such as β_2-microglobulin (B2M) or Serum amyloid A (SAA). Most amyloidoses are linked to aging, and imply extracellular accumulation of precursor proteins, many involve a defective proteolytic process while some are based on intact precursor proteins. Precursor proteins are usually of small size (3-30kDa), are polyanionic, and often rich in aspartic and glutamic acids.

The first step in the formation of amyloid fibrils is the association of precursor proteins into filaments with a beta-pleated structure. In a dynamic process, filaments are assembled into fibrils, which later on form bundles in association with extracellular medium components, and finally fibrous deposits in tissues. It is of importance that all amyloid deposits collected in tissues are not only composed of protein amyloid fibrils, but also of different extracellular components, such as serum amyloid P component or heparan proteoglycans. The study of amyloids in vitro is mainly focused on the fibrillar part of the deposits [157]. A brief description of a few examples of amyloidoses

can help grasping the major characteristics of this class of misfolding diseases.

2.1.1 Alzheimer's Disease

Alzheimer's Disease is a neuro-degenerative disease, often associated to aging, as it is the most common form of senile dementia. It was named after Dr. Alzheimer, who described this disease in 1906 [12]. Symptoms at early stages include inability to acquire new memories, while at later stages, symptoms evolve to mood swings, long term memory loss, confusion, language breakdown. Failure to recognize close relatives, motor difficulties, delusion and resistance to caregiving often imply a lot of stress for the relatives of Alzheimer's disease patients, that are also often the caregivers [166]. The need of healthcare professionals for long-term care of these patients, combined with the high prevalence of the disease (5-8% of population above 65 years, up to 30% above 85), makes of Alzheimer's disease the most costly aging disease for health care systems in developed countries [23], with costs increasing with the age average of the population and age expectancy.

Causes of Alzheimer's disease are still widely debated. While from the 19th century until into the 1960s a vascular cause was accepted for the cerebral atrophy evidence of which was obtained during autopsies of patients with senile dementia, this hypothesis has lost support since then. As soon as 1907 was shown that fibrillar deposits were present in the brain of deceased patients, either inside or outside of neurons [4]. The "amyloid cascade hypothesis" was formulated in 1991 that states that these amyloid deposits were the primary event of the molecular pathogenesis [80], when a pathogenic mutation of the beta-amyloid precursor protein was discovered. This mutation leads to an irregular cleavage of the precursor protein into beta amyloid peptides, that can form amyloid fibrils. This was supported by the fact that Down syndrome patients, who have a third copy of the beta-amyloid precursor protein gene, almost all display Alzheimer's disease by the age of 40 [82]. Another hypothesis is the "tau and tangles hypothesis", that states that the primacy leans toward the formation of neurofibrillary tangles by the tau protein [119], and that amyloid deposits would only be a consequence. The involvement of amyloid fibrils in this disease, either as a cause or a symptom, caused a boost to amyloid fibril formation research, along with the idea that it may lead to the discovery of a treatment. A vaccine was developed, and was successful in preventing the formation of fibrils, but not the development of the dementia, which lead to another toxicity theory for the Abeta peptides, involving protein oligomers (see 2.4).

2.1.2 Creutzfeld Jacob disease and prion protein diseases

Creutzfeld Jacob disease is a transmissible spongiform encephalopathy, a neurodegenerative disease. It was brought to common knowledge during the "mad cow" epizootic,

when it was thought that it could be transmitted via beef infected by the "mad cow" disease (bovine spongiform encephalopathy BSE). This information led to political decisions like the ban of beef products from the United Kingdom inside the rest of the European union from 1996 to 2006, while the british too permissive regulations on cattle feed were blamed for the development of the epizootic. Another possible transmission (which led to another political scandal), was through the injection of Human Growth Hormone obtained from the pituitary gland of CJD dead patients. In both cases, a prion is involved. Prion is a word coined from protein and infection by the 1997 Nobel prize Stanley Prusiner [131]. A prion is an infectious agent made of a protein in a misfolded form, the first identified that doesn't imply a nucleic acid genome. Its key ability is to "transmit" its misfolded state to its well-folded counterparts that are expressed by the infected patient (or animal) [40]. This misfolded state being stable, its propagation is exponential [115]. This prion has a strong propensity to form amyloid fibrils, that leads to the formation of deposits inside the brain and to neuronal death [47]. Other diseases are linked to prions, such as Scrapie in sheeps, Chronic Wasting Disease in elks and mooses, and Feline spongiform encephalopathy in cats.

2.1.3 Other amyloidoses

- Parkinson's disease

Parkinson's disease is also a neurodegenerative disease. It was first described by doctor Parkinson in 1817 [125]. It involves in early stages movement-related symptoms, such as shaking, rigidity, difficulties with walking, while later stages imply cognitive and behavioral symptoms, often leading to dementia. Parkinson's disease is linked to the formation of Lewy bodies inside neurons, that lead to neuronal death. Lewy bodies are aggregates formed around alpha-synuclein amyloid fibrils [158].

- Type II diabetes

Type II diabetes is a metabolic disorder, characterized by insulin resistance which can be combined with lower insulin production. This disease has mainly lifestyle factors, although genetic factors are also known. Insulin-producing islet beta cells are located in the pancreas, and in type II diabetes patients, amyloid fibrils made of the islet amyloid polypeptide (IAPP or amylin) are found in these tissues. IAPP is a small (37 residues) hormone, produced at the same time as insulin. IAPP amyloid fibrils have been proved to be toxic for pancreatic islet beta cells [110]. It has been shown that the pathway of IAPP fibril formation implies the formation of oligomers, especially dimers [51]

- Prolactinoma

Prolactinoma is an adenoma (benign tumor) of the pituitary gland, that implies increased production of prolactin, an hormone similar in structure to growth hormones, which is primarily associated with lactation. Prolactinomas display the formation of prolactin-derived amyloid fibrils, especially linked to aging ([15], [177]).

In conclusion, the study of these various amyloidoses shows us the variety of proteins involved, and the large number of patients that can be affected by these diseases. Especially linked to aging, amyloidoses will see the number of affected patients rise along the increase of life expectancy, making of the study of amyloid fibril formation a more and more important topic of research.

2.2 Amyloid fibrils

The amyloid term was coined in 1838 by Schleiden to describe the amylaceous (starchy) constituent of plants. It was first used in a medicinal context by Virchow in 1851, to describe deposits he discovered in the extracellular medium inside the liver during autopsies, that he thought were made of cellulose. The methods he used to characterize the deposits, based on sulfuric acid and iodine, were indeed used at that time to detect starch. Friedrich showed in 1859 that the main content of amyloids is protein. Congo Red staining was introduced in 1922 by Bennhold, and the use of polarization microscopy in 1927 by Divry, showing that amyloids have a highly ordered structure. Viewed with polarized optics, congo red stained amyloid deposits display a strongly increased birefringence, which is still today the standard characterization test of amyloids. This strong increase in birefringence, that is not seen in other fibrillar structures such as collagen or fibrin, shows that the dye molecules are aligned in parallel along the axis of the amyloid structure ([136], [73],[70]). In 1959, electron microscopy studies showed that amyloid deposits were composed of fibrillar material [39].

Amyloid fibrils are resistant to proteolytic digestion [160] and insoluble in physiological conditions [53]. Amyloid fibrils share a common structure: each fibril (∼7-10 nm wide) contains at least two filamentous sub-units (∼3nm wide) called proto-filaments or filaments, twisting around each other along the fibril axis. The filamentous sub-units are formed from protein polymerized in a cross beta sheet conformation ([22], [53], [71]). These fibrils can also associate with other fibrils, twisting around each other just as the filamentous sub-units they are made of, to form even larger fibrils ([152],[171]), bundles and aggregates [130], or highly organized species known as spherulites [96]. The beta sheets are directed perpendicular to the axis of the filament [53], as shown by Xray diffraction patterns, and infrared spectroscopy showed that these beta sheets are anti-parallel [165].

Amyloid fibrils can be obtained by isolation from tissues, or by in-vitro synthesis. The commonly used protocol to isolate amyloid fibrils from tissues [130] implies treatments

to wash the tissue from soluble proteins and salts, before re-suspension in distilled water to separate the fibrils from other insoluble proteins. Because some amyloidoses are systemic, fibrils can be found in a lot of different tissues, but brain, liver and joints are the most used in the isolation of amyloid fibrils.

2.2.1 Amyloid fibril synthesis in vitro

Amyloid fibrils were first obtained in vitro in 1971 [72], using Bence-Jones proteins. Bence-Jones proteins are immunoglobulin light chain proteins, associated with amyloid light chain (AL) amyloidosis, which is the most common form of systemic amyloidoses [66]. In-vitro fibrils were obtained after 2- to 5-hour incubation in a 50mM guanidine-HCl buffer with acidic pH (pH=3.5), at 37°C, at a concentration of 10mg/mL, in presence of pepsin, which is a proteolytic enzyme that degrades proteins into small peptides. One can already find in this paper major conditions that have been used since then to produce amyloid fibrils: acidic pH, denaturing buffer (even if here the concentration of guanidine-HCl is too low to induce a real denaturation), high protein concentration. It was later shown that the use of guanidine-HCl was not necessary with digested Bence-Jones proteins, as 100mM sodium-acetate buffer could be used just as effectively[107]. On the contrary, a too acidic pH, below 3.6, has been shown to greatly reduce the rate of formation of the fibrils, this being explained by the loss of tertiary structure that occurs at such pH [42].

Insulin can form fibrils in aqueous solution with or without acidic conditions (50mMHCl), at 20 or 85°C, even though the heat treatment largely increases the rate of formation of the fibrils ([126], [99]), proving that intact proteins can form amyloid fibrils, and not only proteolytic fragments. Islet amyloid polypeptide, or amylin, a pancreatic hormone, forms amyloid fibrils both at acidic and neutral pH, a formation that is increased in presence of insulin, which is of relevance as both are involved in type II diabetes and produced in pancreatic beta cells [32]. Fibril in vitro formation protocols can be as simple as incubation in water at 37°C for 30 minutes, as shown for peptides homologous to Alzheimer's disease beta-protein [30], or a saline solution (100mM NaCl), as shown for fragments of the same beta-protein [74]. Amyloid fibrils of B2M were also obtained in an in-vitro cell culture, showing that fibrillation inside a culture of a cell-line coming from patients was possible spontaneously even in-vitro [28]. Polymerization can also be started by the addition of seeds, made of fragments of amyloid fibrils, that act as nucleation center [132]. Fibril formation can be induced by the addition of an enzyme (such as a protease, see Bence Jones protein fibrils) or another biological molecule found in vivo such as a phospholipid, as Lysophosphatidic acid (LPA) was shown to induce in-physiological conditions fibril formation of B2M [129], by destabilizing the tertiary structure of the protein.

To sum up, in a more general perspective, amyloid fibrils are generally obtained in

vitro using slightly destabilizing buffer conditions, like enzymatic, acidic, or organic perturbations. Agitation can be employed, but it has been suggested that the effect of agitation is mainly kinetic, hypothetically creating nucleation seeds from the shearing of the early-formed fibrils that would accelerate the polymerization ([132], [2]).

2.2.2 Amyloid fibril formation by non disease related proteins

Amyloid fibril formation was first studied in vitro by isolation from tissues, coming from patients, and therefore linked to a pathology. After isolation, amyloid fibrils could be studied chemically and later physically. Once their constituent was identified, in vitro studies of the process of formation of these fibrils could be done. As shown during the listing of different pathologies linked to amyloid fibrils, the constitutive proteins or peptides are structurally rather different, some having a alpha-helix-rich structure while others display a beta-sheet-rich one, even though the structure of fibrils being a beta-sheet stacking, all share a beta-sheet conformation propensity. In 1998, amyloid fibrils were obtained, which were constituted of a protein (SH3) not related to a pathology. Conditions were similar to what was previously used to obtain amyloid fibrils in vitro, meaning incubation in a phosphate buffer at 4°C or room temperature for a few days, at pH=2. Amyloid fibrils were totally identical in structure to pathology-related fibrils, showed a birefringence increase under Congo red staining, fibrillar aspect in electron microscopy, and a cross-beta structure in Xray diffraction. This study showed that the amyloid fibril formation could be accessible to proteins that were not linked to a pathology, and that the key element was to have access to destabilized states of the protein [79]. Since then, many other proteins, not related to amyloidoses, have been shown to be able to form amyloid fibrils, such as muscle myoglobin [62], a cold shock protein [78] or a helical cytochrome [127] for example.

2.2.3 Amyloids' functional role

Just as was seen in vitro, amyloid fibrils not linked to diseases were found in vivo [38]. So-called "functional" amyloids have been used by living organisms with different roles, at one point or another of their life-cycle. For example, E. coli uses amyloid fibrils, made of the protein "curlin", to colonize inert surfaces, and to bind to different host proteins [31]. These functional amyloids can also be found in mammalian systems, as it was observed that fibrous striations involved in the production of melanin are composed of amyloid fibrils [14]. Functional amyloids are exactly similar in structure to disease related amyloids, as they show birefringence upon Congo red staining, have the same diameter as seen in electron microscopy, show an extensive beta sheet structure as seen in Circular Dichroism (CD) [63]. In a similar manner, the heritable conformational changes of prion proteins have in some cases been proved beneficial,

and inheritable as a non-Mendelian non-chromosomal genetic trait. For example the polymerization of Ure2p protein into amyloid fibrils results in the activation of genes involved in the uptake of poor nitrogen sources [35]. The polymerization of the CPEB (cytoplasmic polyadenylation element-binding) protein from a marine snail has been suggested to be essential for the long-term maintenance of synaptic changes associated with memory storage [153]

2.3 Amyloid fibril formation mechanism

The formation of amyloid fibrils through a nucleated growth mechanism is widely accepted [151]. The mechanism starts with a lag phase, during which the formation of fibrils is not significant, as established using Thioflavin-T (ThT) fluorescence or light scattering, among other techniques. During this lag phase, different oligomeric constructs form, some of which have a beta-sheet-rich structure and will serve as polymerization nuclei. The use of polymerization seeds, made from the shearing of regular fibrils, induce the disappearing of this lag-phase, because the polymerization nuclei are provided. The lag phase is then followed by an exponential growth of the fibrils[121]. Many studies have been focusing, in the past decade, on these oligomeric, pre-fibrillar constructs, both for the fact that they may be the toxic agent in amyloidoses, and that to understand completely the mechanism of formation of amyloid fibrils, the formation of the polymerization nuclei can't be avoided.

For many proteins (including B2M and Abeta, responsible for Alzheimer's), these oligomers have been identified, using either atomic force microscopy (AFM) or Transmission electron microscopy (TEM) . Some of them appear as spherical beads (2-5nm in diameter), or linear chains made of these beads, and even annular structure, that could be formed by the circularization of these beaded chains [81]. They have also been recently characterized by Small Angle Xray Scattering (SAXS), showing beaded-chains of insulin, on-pathway of the formation of fibrils [171]. These structures were termed protofibrils by their first descriptors, and are not to be confused with protofilaments. Protofibrils can bind ThT and show birefringence dyed with Congo Red, and have a beta-sheet-rich structure [173]. Protofibrils made from different proteins share structural elements, as a protofibril-specific antibody can bind all of them but not the corresponding monomers or mature fibrils [98]. Protofibrils themselves are preceded by low molecular weight oligomers, dimers to hexamers for Abeta, that from CD experiments seem to be rather disorganized [16]. Interestingly, these low molecular weight oligomers are found in the brain of Alzheimer's disease patients [17]. These low molecular weight oligomers have been identified for Sup35p yeast prion [151], insulin [171] and the SH3 domain from a bovine kinase [7], among others.

2.4 Amyloids and oligomers: paradigm shift for toxicity

The discovery of fibrillar deposits in the organs of patients suffering from certain diseases made reasonable the assumption that amyloid fibril deposits were the cause of these diseases. Moreover, the injection of amyloid fibrils into the cerebral cortex of aged monkeys provoked neuronal loss [67]. But recent findings have shown that amyloid fibrils may be a symptom, and not the cause of the diseases, while fibril-precursors, such as oligomers or proto-fibrils, would be the pathogenic agent, at least in neuropathic diseases. To limit evidence to Alzheimer's disease, it has been shown that the severity of the cognitive impairment correlates with the levels of oligomers, but not that of fibrils [111]. In transgenic mice, the same symptom appears well before the accumulation of amyloid deposits in the brain [117]. And at last, a mutation responsible for a heritable early onset manifestation of the disease is proved to provoke oligomerization but no amyloid fibril formation [122]. These findings are corroborated by the toxicity of prefibrillar forms of non disease-related proteins, while their native or fibrillar form are not toxic [38]. The enhanced toxicity of low molecular-weight oligomers compared to amyloid fibrils may be explained by the fact that smaller aggregates have a higher proportion of residues on their surface, residues that in properly folded proteins would not be exposed. These exposed residues may be involved in aberrant interactions with the cellular machinery, causing its dysfunction.

Chapter 3

Experimental techniques

Most biophysical techniques have been used to study folding or amyloid fibril formation, with either structural or kinetic focuses, and those that are mainly used are briefly described in this part. The study of these phenomenons is the study of complex samples, involving the existence of transient species, polydispersity, aggregation propensity, different time and size scales, and fast kinetics, making the use of complementary techniques a method of choice.

3.1 UV-visible absorption spectroscopy

UV-visible (UV-vis) spectroscopy is probably the most widely used technique for routine analytical work, such as protein detection and concentration measurement, along with rough protein characterization [154]. The method is based on the interaction of light, inside visible (400-700nm) and UV (200-400nm) wavelength ranges, and electrons of for example delocalized π-systems. Protein concentration can be measured with absorption at 280nm, using Beer's law and calculated (or measured) molar extinction coefficients. UV-vis spectroscopy can also be used to follow molecular interactions or changes in the structure of proteins. Aromatic amino-acids are powerful probes in the ultraviolet spectral absorption range. The spectrum of a chromophore is highly dependent on the polarity of its microenvironment. An aromatic residue will have a different spectrum depending on whether it is exposed to solvent or buried inside the protein for example.

Using such a property, folding and unfolding (denaturation) of proteins can be followed. Denaturation can be caused by successive adjuncts of urea or guanidinium, or pH, ionic strenght and temperature changes. Folding kinetics can be followed using denaturant dilution or pH jumps. If aromatic residues, such as tyrosine or tryptophan for example, are involved, interactions can be detected, such as protein-protein

or enzyme-substrate binding or protein-metal interactions. Artificial chromophores can also be attached to specific parts of the protein, through mutation of a residue to an aromatic one or through disulfide linkeage to a cysteine, providing probes in regions that can't be detected with this technique due to the absence of natural chromophores.

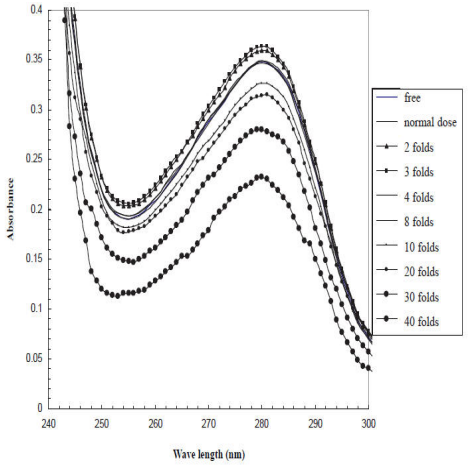


Figure 1. UV-visible spectrum of HSA in the presence of different doses of co-amoxiclav in the phosphate buffer 50 mM, pH 7.5, at 37°C.

Figure 3.1: Evolution of the human serum albumin (HSA) UV-vis spectrum in presence of growing concentrations of a drug. At high concentration, HSA is denatured by the drug. [164]

3.2 Spectrofluorimetry

Fluorescence is a light emission phenomenon, implying a transition from a higher to a lower energy state measured by the detection of the emitted radiation rather than the absorption as seen in spectro UV-vis. In order for this transition to occur, the molecule must have been excited, for example through the absorption of light. Different wavelengths in absorption and emission are due to energy losses as heat, during collisions with other molecules, called the Stokes shift (see figure 3.2). Most molecules that absorb in the ultraviolet/visible region of the spectrum do not fluoresce, as their excited electrons go back to ground energy state through heat emission, but many fluorescent compounds are of biological interest.

The technique of spectrofluorimetry is most accurate at very low concentrations, whereas absorption spectrophotometry is least accurate under those conditions. Fluorescence is highly dependent on pH, solvent composition, and temperature [87]. This technique can also be used for quantitative determination of low-concentration compounds, for example vitamin B1 in foodstuffs [77]. Fluorescent probes can be used to detect metal ions such as Ca^{2+}, or can be the product of an enzymatic reaction [26].

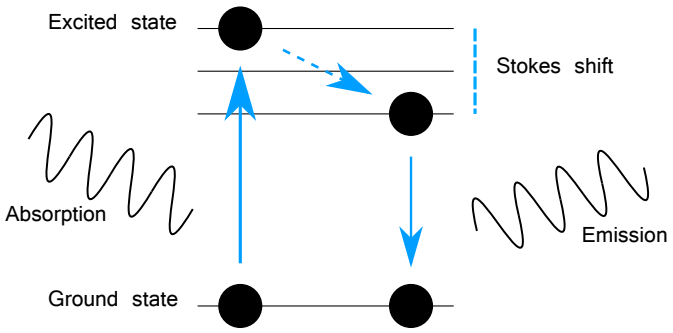


Figure 3.2: The Stokes shift explains the difference between the wavelengths of absorption and emission

Redox reactions involving NADH and NADPH can be followed kinetically in vitro at in-vivo similar concentrations, or even in intact cells or mitochondria. Changes in protein structure, using either FAD as tryptophan cofactor or a chimera with GFP, can induce fluorescence changes as conformation or microenvironment around the probe changes. Thioflavin T is a widely used probe, as it is specific to amyloid fibrils (Figure 3.3).

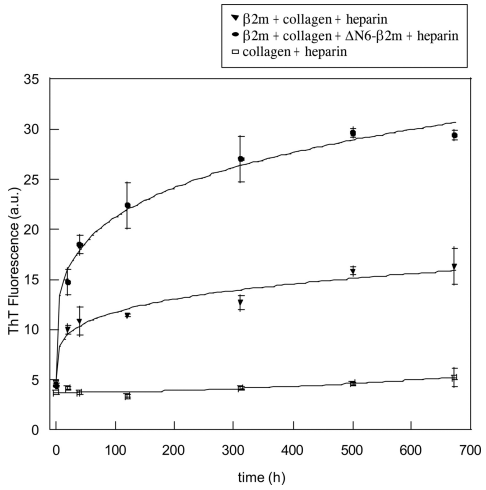


Figure 3.3: Formation of amyloid fibrils of B2M in presence of several co-factors, followed using Thioflavin-T fluorescence.

3.3 Circular dichroism

Electromagnetic radiation (light) oscillates in all possible directions, but it is possible to select preferentially waves oscillating in a single plane using polarising material. Polarimetry is essentially the measure of the angle through which the polarisation plane is changed after passing through a solution containing a chiral substance. In

circular dichroism, incident light is circularly polarized (composed of two 90° phase-shifted perpendicularly polarized waves with identical intensities), and undergoes a different interaction with chiral compounds on different polarization planes, inducing an elliptically polarized resultant beam. This ellipticity is measured in CD experiments and is plotted against wavelength, giving information on secondary and tertiary structure. The CD spectrum of a protein is compared, through curve-fitting, to those of poly-L-amino acids, that have known alpha helix, beta sheet, and random coil contents, and a percentage of the content in these secondary structures in the protein can be estimated (figure 3.4 [86]). Using this estimation, the acquisition of secondary structure during folding or unfolding can be easily followed, giving access to thermodynamic and kinetics parameters [101]. Another example is the determination of binding constants in protein-protein or protein-substrate interactions[114]. Tertiary structure changes can also be followed, through the burying or exposition of tyrosine and tryptophan residues, using far UV wavelength.

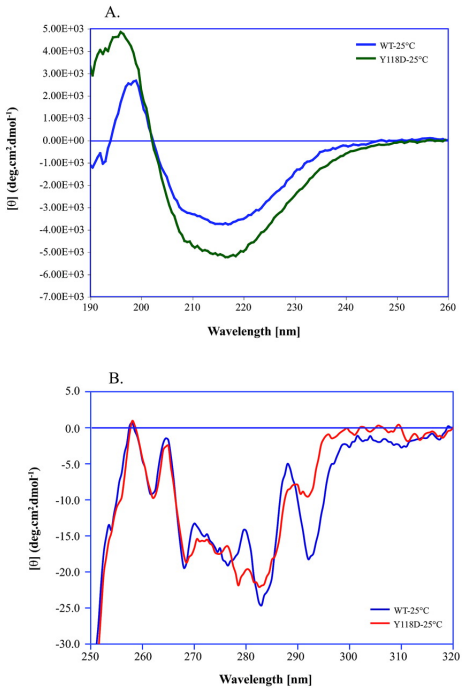


Figure 3.4: Near (A) and far (B) circular dichroism spectra of alpha A crystallin and its mutant Y118D. Differences are consistent with an increased alpha helix content in the mutant protein.

3.4 Infrared spectroscopy

Below the UV-vis energy range in the electromagnetic spectrum is the infrared region (700nm to 25µm). The absorption energy corresponding to this level is that of

vibrational states transitions. Bond between atoms can be considered here as flexible springs, allowing the molecule to adopt several vibrational modes, including stretching or bending. A molecule can absorb incident light by switching to the specifically higher energy vibrational mode. An infrared spectrum is thus an absorption spectrum. As the number of functional groups increases, the molecules becoming more complex, infrared absorption bands become more difficult to assign, due to the overlay of bands close in frequency, but also due to the influence of the molecular environment on the frequency itself. For example it is possible to distinguish between the C-H vibrations of CH_2 and CH_3.

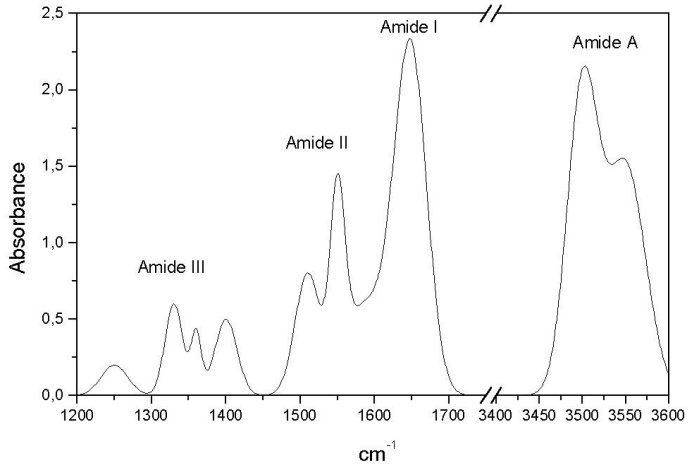


Figure 3.5: Typical protein infrared spectrum with assignment of the amide bands [90]

While infrared spectroscopy is mainly used in chemical and biochemical research of small compounds such as drugs or substrates, and can be coupled to a separating method such as gas chromatography, it can be part of protein analysis. Especially, the peptide bonds gives rise to nine absorption bands, named amide A, B and I through VII, amide I and II being the major contributors. Both bands being related to the backbone conformation, they are characteristic of secondary structure [9]. Secondary structure changes can be followed using time-resolved infrared spectroscopy, specific evolutions of the bands being linked to transitions in the alpha helix, beta sheet or random coil content. In practical terms, transmission fourier transform infrared spectroscopy (FT-IR) is highly contaminated by the strong absorption of the O-H bond vibration, so liquid-state experiments on proteins have to be conducted in D_2O, with questions arising on the biological relevance of the solvent, and on the fact that folding processes can be different in D_2O and H_2O. To tackle this problem, attenuated total reflectance (ATR) spectrometers can be used. ATR uses evanescent waves to measure the spectrum of a sample deposited on a measurement crystal. Evanescent waves are absorbed inside the sample in a layer which thickness is of the order of

the wavelength, circa 1-10µm. This means that the optical thickness of the sample is compatible with measurements in H_2O.

Fourier transform infrared spectroscopy (either transmission or ATR) allows, through the use of interferograms, measurement times of the order of seconds, instead of tens of minutes for the classical dispersive infrared spectroscopy.

3.5 Small angle X-Ray Scattering (SAXS)

SAXS (and Small angle Neutron scattering where neutron beams replace X-rays, with very similar principles) is a technique that allows the determination of size and shape parameters of large molecules such as proteins. X-rays are diffracted by electrons, and as such analysis of X-ray crystallography diffraction data sets produces an electron density map of the crystal. But contrary to X-Ray crystallography, where a bias is introduced as molecules are trapped in a crystal, in SAXS the proteins are in aqueous solution. Due to the random positions and orientations of molecules in the liquid state, the intensity distribution is isotropic, and proportional to the scattering intensity of one molecule averaged over all orientations. Experimentally, the scattered intensity is recorded as a function of momentum transfer q ($q = \frac{4\pi sin\theta}{\lambda}$), where 2ϑ is the angle between the incident and scattered radiation. At small angles, the scattering curve is a rapidly decaying function of q, essentially determined by the shape of the molecules in the sample. Guinier analysis (equation (3.2)) provides estimations of the radius of gyration of the molecule and the extrapolated forward intensity, which is proportional to the molecular weight of the molecule, which can be obtained by comparison to a standard (equation (3.3) where C is the massic concentration). Fourier transform of the scattering intensity curve provides the pair distribution of the molecule, which is a histogram of interatomic distances. Comparison to the pair distribution of standard shapes provides hints of the shape of the molecules in solution. Using simulation of scattering curves and fitting them to the experimental curves, softwares can calculate a rough three dimensional structure, or an ensemble of structures of the molecule, with a resolution up to 10Å. For a detailed theoretical approach of SAXS, see [162].

$$I(q) \approx I(0) \exp\left(-\frac{1}{3} R_g^2 q^2\right) \tag{3.1}$$

$$\ln\left[I(q)\right] \approx \ln\left[I(0)\right] - \frac{1}{3} R_g^2 q^2 \tag{3.2}$$

$$I_{protein}(0) = I_{standard}(0) \frac{C_{protein} M_{protein}}{C_{standard} M_{standard}} \tag{3.3}$$

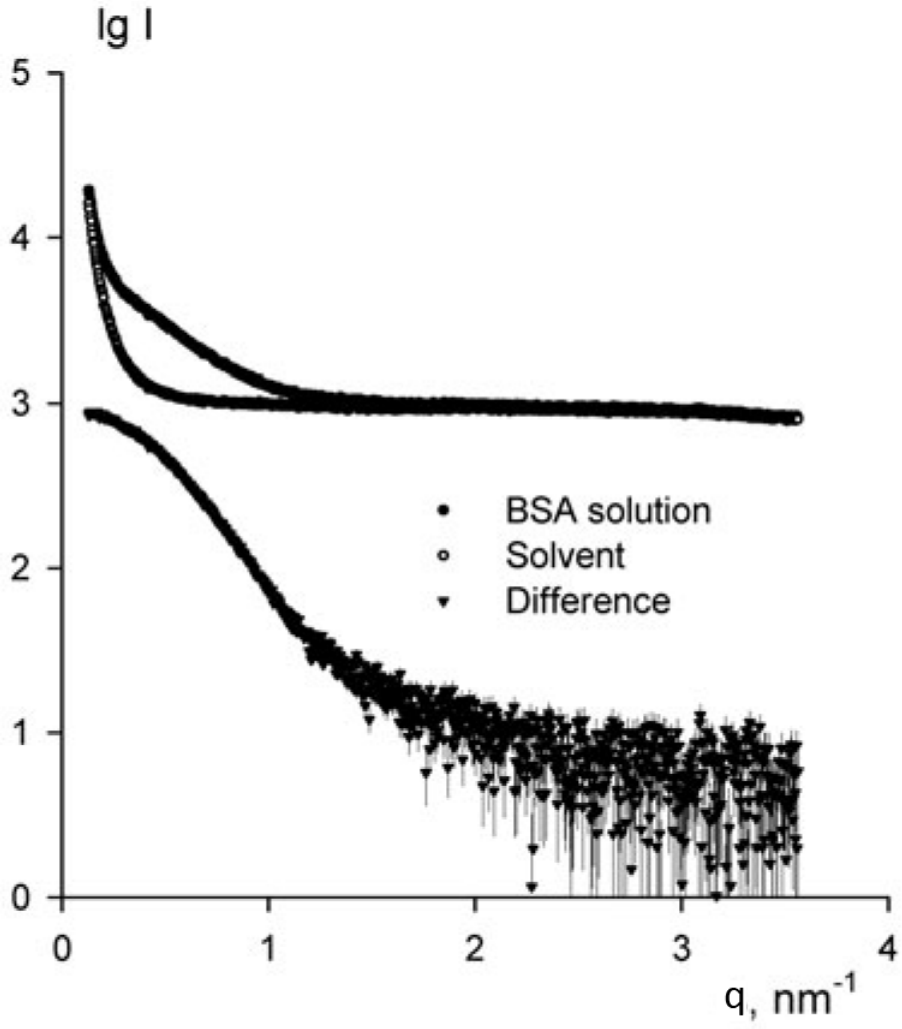


Figure 3.6: Typical scattering curves of BSA before and after subtraction of the solvent scattering curve.

Standard conditions for successful SAXS experiments imply a monodisperse macromolecule solution. Indeed as soon as the sample is not monodisperse, scattering curves measured are the sum of scattering curves of the different compounds in the solution. Of course calculating the "average-shape" of a mix of proteins makes no sense. But one can follow the formation of a dimer, or a protein complex, by recording the increase of the average gyration ratio or average molecular weight inside the sample. The presence of oligomers can be proven, and the evolution of the oligomerization state of a protein can be followed. Several tools have been recently developed, such as singular value decomposition, to help in the analysis of polydisperse samples. Singular Value Decomposition is a mathematical data analysis that derives from the scattering curves the singular vectors and values that compose the signal. It can be an indication of the number of compounds present or of states that are populated, or even the mechanism of formation of amyloid fibrils [171].

3.6 Transmission Electron microscopy

While light microscopy is limited to a routine resolution of 0.5µm, electron microscopy can provide a resolution up to 1nm. In electron microscopy, the light source is replaced with an electron gun, consisting of a high voltage passed between the cathode (a tungsten filament) and the anode in which a hole allows the formation of the electron beam. While passing through the sample, some of the electrons of the beam are scattered, while others are then focused on a digital camera. When the technique was developed, samples had to be stained with high atomic number compounds, like heavy metals such as uranium, lead or osmium, making labelled structures appear dark or electron dense.

Negative staining techniques on the contrary offer the possibility of contrast enhancement, with faster and simpler protocols. The idea of negative staining is to surround or embed the object of interest with an electron dense material that provides high contrast. Instead of appearing dark on a light background, the biological object will appear light on a dark background. These techniques have been used, with different stains, since the 1950s, and allowed the collection of information on the structure of ribosomes, viruses, bacteriophages, microtubules, and amyloid fibrils and other biological filaments.

3.7 NMR spectroscopy

Similar to other spectroscopic techniques, NMR implies an energy input to promote the transition from a ground state to a higher energy state, specifically here nuclear

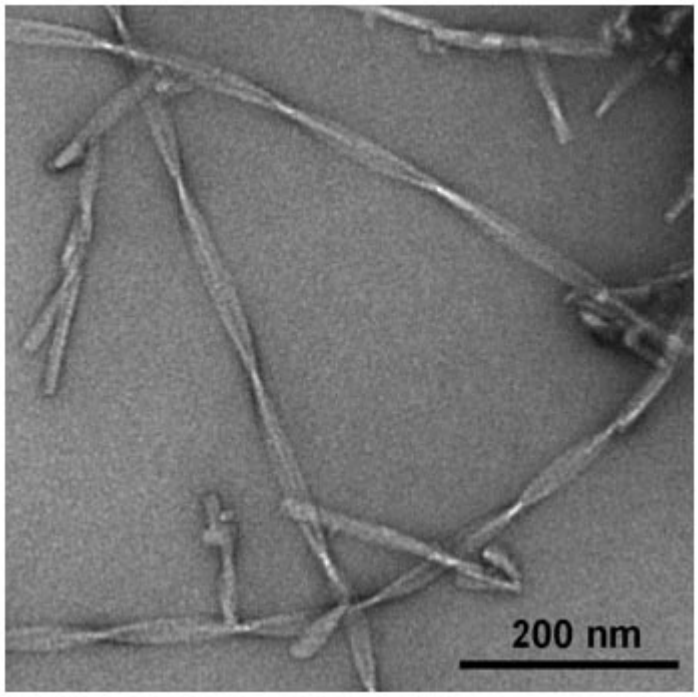


Figure 3.7: Amyloid fibrils shown using negative staining electron microscopy, from [142]

magnetic spins. Nuclear magnetic spins are intrinsic properties of nuclei which populate energy levels according to quantum theory. After a certain time span, spins excited by an electromagnetic radiation relax to the ground state, fixed by exposition to a high intensity magnetic field. The spins resonate at a frequency linked to the value of the magnetic field produced by the coil of the superconductive magnet, but shifted by a value linked to micro-environmental magnetic fields produced by adjacent nuclear spins. This frequency shift is termed chemical shift, and is the most common parameter measured by NMR. Along with common ^{1}H NMR spectroscopy, the study of bio-molecules almost always involves ^{13}C and ^{15}N spins. Because of the low natural abundance of such isotopes, recombinant proteins and nucleic acids are usually labelled during their production using isotope-enriched bacterial nutrients (^{13}C glucose, ^{15}N amino-acids) . To reduce spectral complexity,^{1}H atoms can be replaced, during production or through exchange, by ^{2}H isotopes.

- Multidimensional NMR

Common NMR spectra are obtained using a pulse acquired Fourier transform scheme, where the observable is the decay of the transverse magnetisation, called free induction decay (FID). The detected signal is thus a function of the detection time t_2. Protein spectra usually show a strong overlap of signals, as shown on figure 3.8 (top), due

to the high number of spins that are present and detected within the sample. To tackle this difficulty, the spectra can be "spread" on a second dimension or more. For example, the position of peaks in a 2D spectra can correspond to the chemical shifts of both an hydrogen atoms (x axis) and its bonded nitrogen (y axis). This is obtained using another time t_1 (evolution time), that can be introduced in the pulse sequence at specific positions, and the signal becomes a function of both t_1 and t_2, while its Fourier transform comprises two frequency components. An example is shown on figure 3.8 (bottom).

NMR techniques specifically used for our studies along with theoretical principles are further described in chapter 7.

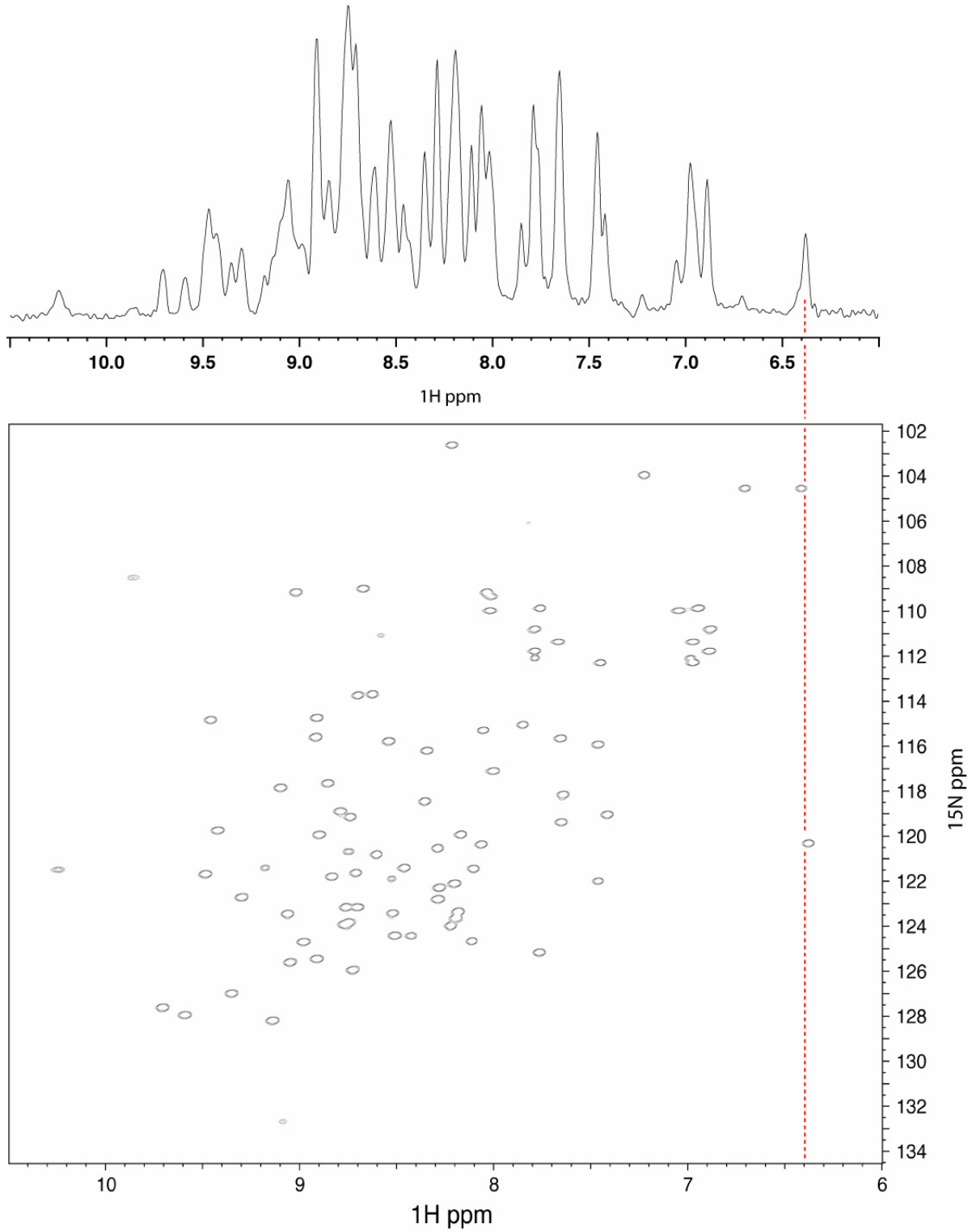


Figure 3.8: Typical NMR spectrum of the amide region of a protein. The 1D spectrum (top) shows large overlaps (degeneracy), which can be much reduced using Nitrogen frequencies in an indirect dimension on a 2D spectrum (bottom).

Part II

β_2-microglobulin: a model protein for amyloid formation and protein folding

Chapter 4

B2M: general introduction

β_2-microglobulin (B2M) is a protein that has been used as a model system for many years, due to three main characteristics. First, B2M forms amyloid fibrils, in vivo and in vitro. Second, these amyloid fibrils are involved in a disease called Dialysis related amyloidosis (DRA), giving its study a direct medical relevance. And finally, while small proteins usually fold within a few milliseconds at most, B2M folds within a dozen minutes in physiological condition, and is as such a good model to study protein folding.

4.1 Isolation, structure, function

B2M was first isolated in 1968 in urine from patients with tubular proteinurias (malfunction of the renal tubules) [13], and shown to be present in low amounts in serum, urine, and cerebrospinal fluid of sane individuals. It was fully sequenced in 1972 and by homology described as a free immunoglobulin domain [128]. Finally, B2M was linked to its function in 1973, as it was shown to be the small sub-unit of the Human Leukocyte Antigens (HLA) complex[76].

HLA is the human major histocompatibility complex (MHC), and is involved in the recognition of self and non-self, an essential part of the immune system. HLA are present in all nucleated cells, thus excepting red blood cells. Peptides generated by proteolysis inside the cell are presented outside the cell by these MHC antigens, so that they can be identified as self or non self markers by cytotoxic T cells (Leukocytes CD 8). Each MHC molecule can contain only one peptide at a time, although peptide-specificity is very broad, as each MHC molecule can bind many different peptides. MHC molecules acquire their peptide during their own bio-synthesis, inside the Endoplasmic reticulum, and this acquisition is essential for efficient transport of the MHC molecule to the cell surface [167]. This complex has a long lifetime in vitro

(tens of minutes), even longer in-vivo, and a low Kd (4nM), making of it a very stable complex [88].

B2M has an essential role in the binding of peptides by the MHC molecule inside the Endoplasmic reticulum, as B2M-deficient cell lines were not able to present peptides at their surface in a stable manner [172]. Moreover, MHC molecules in solution displayed an enhanced affinity toward peptides in presence of B2M. B2M binding is believed to act as a transition factor from a low affinity to a high affinity state [25]. B2M being a free domain, it can be exchanged with another B2M present in the intra- or extra-cellular medium, which facilitates the exchange of the peptide contained inside the MHC[104]. B2M is also involved in the efficient transport of the MHC molecules from the Endoplasmic reticulum to the cell surface, as was shown using B2M-deficient murine cell lines [178].

MHC class I molecules are composed of two sub-units: one heavy chain, alpha, composed of three sub-domains, and one small chain: B2M. The two chains are exposed to the extra-cellular medium, and attached to the cell membrane via a trans-membrane domain in the alpha3 sub-unit. B2M lies aside this alpha3 sub-unit, while the alpha1 and alpha2 domains, which are very polymorphic in order to bind the different peptides, are paired to form the cleft-like peptide binding site, as shown in figure 4.1.

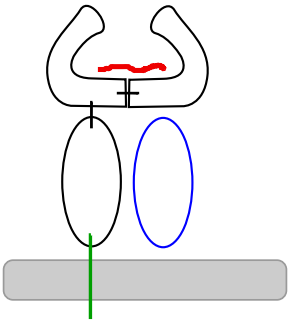


Figure 4.1: Model of a MHC class I heavy chain in complex with B2M. In green the transmembrane domain, in black the three heavy chain subunits, in blue B2M. The peptide, displayed in red, is bound inside a cleft made of two alpha subunits of the heavy chain.

The first structure of the HLA complex was obtained in 1987 [18] from X ray crystallography, and is shown in figure 4.2. This structure displays both alpha and B2M chains. Membrane proximal domains (B2M and alpha3 sub-unit) have an immunoglobulin-like fold. The contact interface between B2M and alpha3 is made of their respective 4-strand beta-sheets. The alpha1 and alpha2 sub-units, each consisting of a 4-strand beta-sheet and helical region, that pair to form a unique 8-strand beta-sheet topped by two helices, forming the cleft-like peptide bonding site. Peptide binding being essential for stability of the MHC complex, this structure, as most structures of the MHC complex [145], [94], has a filled peptide bonding site. Peptides

are usually tightly bound to the cleft from their carboxyl and amino termini, through an interaction between the alpha 1 and 2 domains and the polar atoms from the peptide main chain [112] (figure4.3). The generic interaction allowed by the involvement of the main-chain atoms implies a tight binding for all types of peptides, and a tolerance for peptides longer than the nanomeric peptides that usually bind the MHC molecules [56], [139].

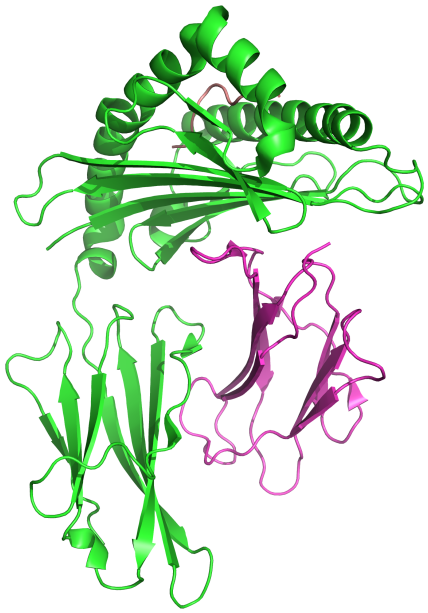
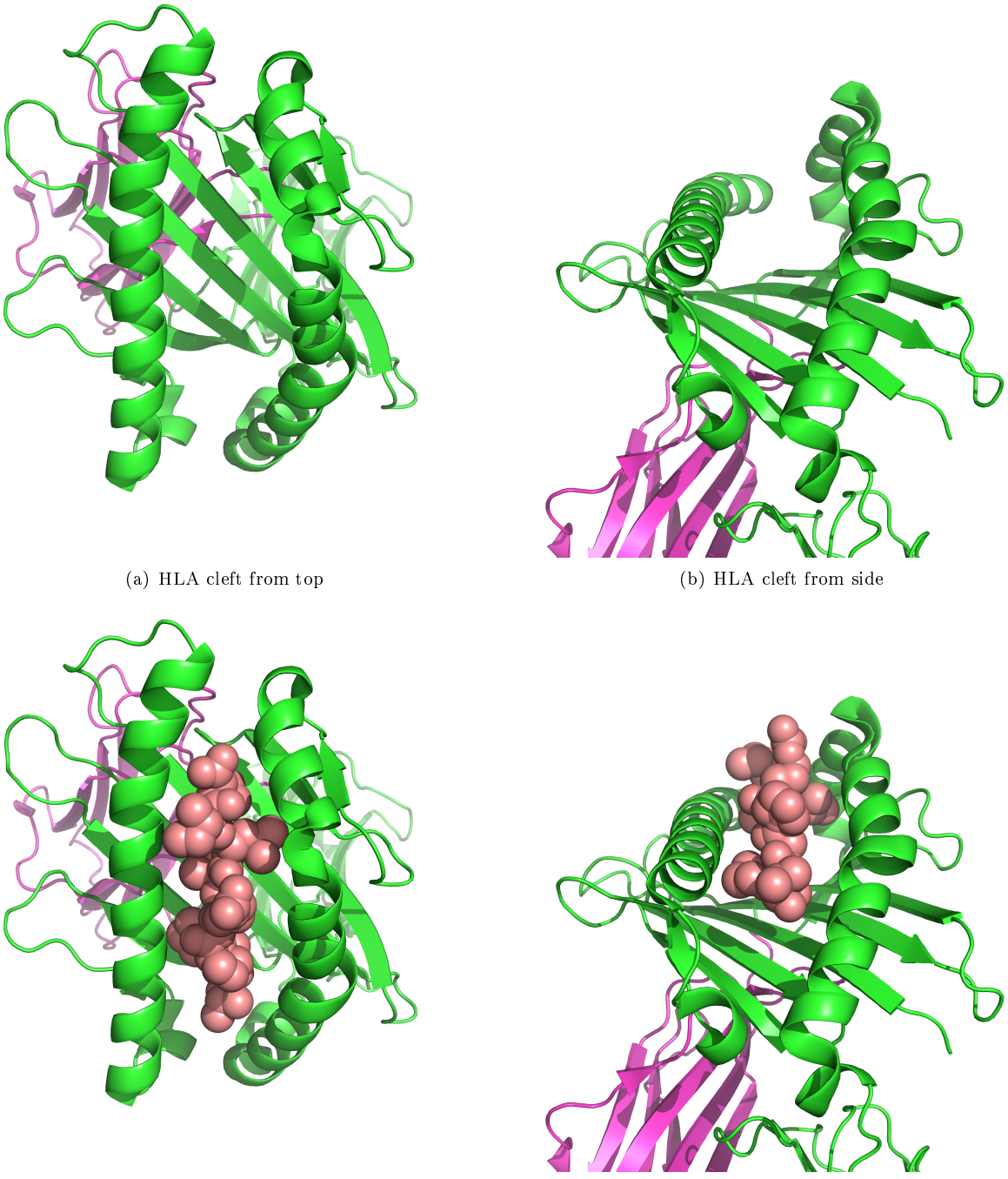


Figure 4.2: Crystal structure of the extra-cellular part of the HLA complex, heavy chain in green, B2M in purple



(a) HLA cleft from top

(b) HLA cleft from side

(c) HLA cleft with peptide from top

(d) HLA cleft with peptide from side

Figure 4.3: Details of the peptide-presenting cleft of the HLA complex

The structure of B2M alone was obtained in 1985 for bovine B2M [10], in 2001 for human B2M [168], [170], while it had been obtained in complex inside the MHC in 1991 [145]. B2M is a 119 residue protein, which includes a 20 residue signal peptide,

leading to a 99 residue secreted protein, with a molecular weight of approximately 11.7kDa. B2M displays the beta-barrel like structure of immunoglobulin domains. It consists of two beta sheets containing respectively three and four beta strands. The two sheets are folded over one another in a sandwich-like conformation, and linked by a central disulfide bond between cysteins 25 and 80, which are the only cysteins of the sequence, as shown on figure 4.4. The beta strands are traditionally named with letters from A to G, according to the primary sequence order, the strands A (6-12) B (22-30) D (51-56) E (62-69) forming one beta-sheet, the strands C (36-41) F (78-83) G(91-94) the other, as shown the secondary structure map of B2M, figure 4.4.

Overlap of different structures, obtained either in complex inside MHC molecules or alone in a crystal or in solution, is very high, with a main-chain root-mean-square deviation (RMSD) of the order of 1 Angstrom. Event though the global fold of B2M is manifestly conserved, as seen in the different structures and conditions, one important specificity of B2M's structure is its flexibility. Both by NMR and crystallography, the D strand, for example, is subject to conformational equilibrium dynamics, while the N terminus of a single residue mutant of B2M populates two dramatically different orientations inside the same crystal [137]. Another example is the existence of a beta bulge inside the beta strand D when in complex inside a MHC molecule, dividing it into two smaller strands, while this strand is fully formed in structures in which B2M is isolated [138]. Flexibility can also be seen in some loops, especially the DE loop, a loop that will be shown of importance when discussing amyloidogenesis of B2M.

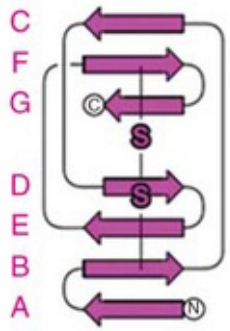
4.2 B2M and Dialysis Related Amyloidosis

Dialysis related amyloidosis is a serious and frequent complication found in patients undergoing long term hemodialysis, that was first reported in 1975, when the hemodialysis technique was already 15 years old [175]. It mainly involves B2M deposition in joints in the form of amyloid fibrils [65]. These deposits induce bone abnormalities, such as carpal tunnel syndrome and destructive arthritis [33]. B2M amyloid fibril formation is linked to a $\sim$30-fold increase of circulating B2M, normally catabolized in the renal tubules, and not efficiently filtered by dialysis membranes [64]. With the improvement of medical techniques, patients undergoing hemodialysis can live a prolonged life, but as prevalence of DRA increases with the duration of the dialysis treatment, reaching almost 100% after 13 years of treatment [91], complications and quality of life reductions are expected to rise as well, as already observed [179]. Progress in dialysis techniques, especially the use of high-flux membranes and specific B2M traps was able to decrease the prevalence of DRA [169], but is not sufficient to emulate the B2M elimination by functional renal tubules. Medical and surgical treatments can be used to relieve patients from symptomatic pain, but none has been found yet to prevent DRA or bone abnormalities. The only preventive treatment is

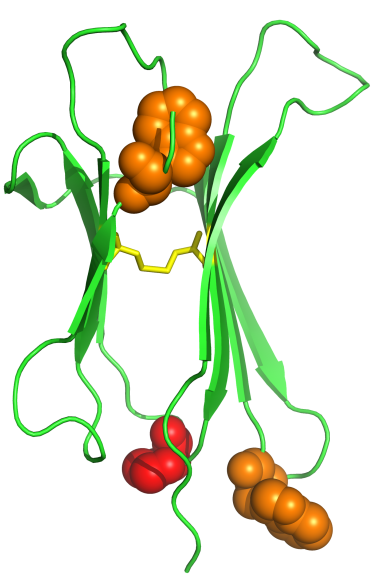
renal transplantation, but it has to be made early as its ability to eliminate amyloid deposits is still controversial. This is why understanding B2M fibril formation and toxicity is a key element in the perspective of finding a medical treatment.

4.3 B2M as a marker for other diseases

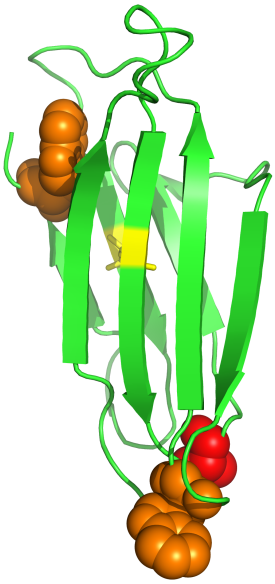
B2M being part of the immune system through participation in the major-histocompatibility complex, its circulating concentration has been shown to be linked to the activation of the immune system. As inflammatory processes take place, the rise of the immune response is linked to a higher production of B2M, and as a consequence a rise in the serum circulating concentration of B2M. B2M has indeed been used as a marker of the development of pathologies, such as sarcoidosis, multiple sclerosis, certain auto-immune diseases, as a tumor evolution marker in certain carcinoma, and even as a marker of the evolution of HIV in certain patients ([106], [1], [3]). It has also been shown to be a good surrogate test for HIV infection for use by insurance companies in jurisdictions where HIV-antibody testing of insurance applicants is forbidden [21], science has its moral failures.



(a)



(b)



(c)

Figure 4.4: (a) secondary structure map; (b) and (c) B2M crystal structure, with the internal disulfide bond in yellow, W60 and W95 in orange, P32 in red. On (c), β-strands A, B, E, D are in front, left to right respectively, while strands C, F and G are in the back, left to right respectively.

Chapter 5

B2M: amyloidogenesis

5.1 B2M discovered as an amyloidogenic protein

B2M has been linked to amyloid fibrils in 1985 from N-terminal sequencing experiments performed on fibrils purified from tissues of a chronic hemodialysis patient ([65], [75]). The same year, amyloid fibrils were obtained in vitro from pure B2M [43], with a very simple protocol: B2M dissolved in 10 mM Phosphate Buffer Saline (PBS) was dialysed against decreasing concentrations of PBS, and finally water, and then left to dry overnight.

Deposition of B2M amyloid fibrils in vivo, in DRA patients, is linked to a concentration of circulating B2M more than 30-fold higher than in healthy patients, as B2M is no longer eliminated in the kidneys. But the correlation between concentration of B2M, either in vivo or in vitro, and amyloid fibril formation or deposition, is poor [64]. For example B2M remains monomeric for months in physiological conditions of pH and temperature, at a concentration more than 20-fold higher than that found in DRA patients [120]. For this reason, other factors have been suggested that may increase B2M fibrillation, such as duration of kidney failure [46], post-translational modifications [29], or interactions with other molecules in vivo [120], or in vitro interaction with Cu^{2+} [118], glycosaminoglycans [120], lysophosphatidic acid [129] or collagen [68].

5.2 B2M fibril formation mechanism

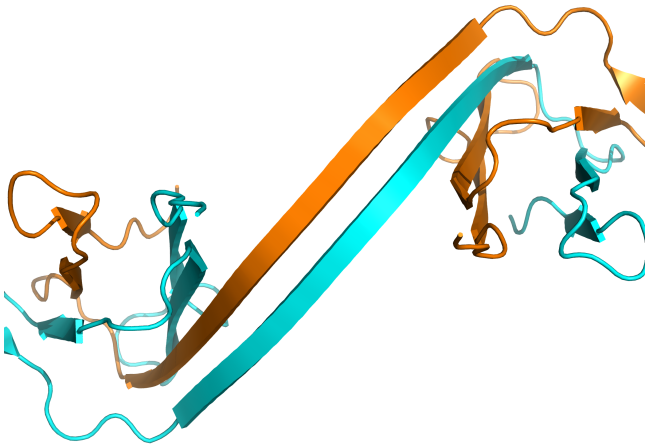
Many experimental set-ups have been used to study the mechanism of B2M fibril formation. Most of them focus on the destabilization of the protein, either by mutation (of Proline 32 or in the DE loop) or truncation (DeltaN6, DeltaK58), biochemical

interaction with Cu^{2+} [118], organic solvents [135] or collagen[68], or physico-chemical constrain such as ultrasonication, heat, high salt concentration and agitation. All those changes, either intrinsic or environmental, have an influence on the cis/trans equilibrium of the His31-Pro32 peptide bond. While it is in a cis conformation in the native state, a trans conformation is observed for this peptide bond in fibrils, as was recently demonstrated using ssNMR [8]. Other regions such as the DE-loop have been highlighted as important for the formation of fibrils, as the perturbation of the dynamics inside this part of the protein can have a dramatic effect on polymerization. For example, full length mutants that are cleaved at lysine 58, have been shown to be highly aggregation prone [83], while the W60G mutation diminishes the capacity of the protein to elongate fibrils.

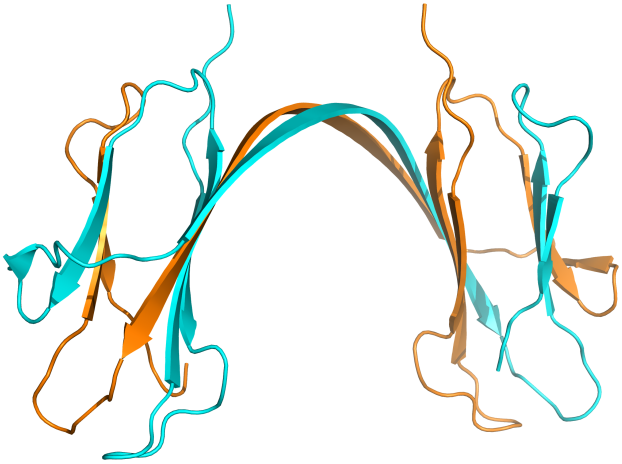
5.3 B2M and oligomers

As for other proteins, and in accordance with the standard model of the amyloid fibril formation mechanism, it is suspected that in-vivo B2M forms oligomeric intermediates on the pathway to mature fibrils [55]. To study this phenomenon, several destabilizing conditions have been used such as copper adjuncts, reductant conditions (DTT), or binding with nanobodies.

In 2011, Liu and coworkers [108] used incubation in presence of DTT to induce fibril formation. They observed that a ladder of oligomers were formed in the process, with all stoichiometries, from dimers to higher than decamers. The resistance of these oligomers to SDS, along with their susceptibility to DTT reduction in the loading buffer used for a SDS gel, indicates that these oligomers are covalently bound through disulfide bridges. The Xray diffraction structure of this dimer reveals, as expected, inter-molecular disulfide bonds, through the rearrangement of the intra-molecular bond that is present in native B2M. More strikingly, the dimer exhibits a domain-swapped structure, as shown on figure 5.1, with identical domains (beta-strands E, F and G) being exchanged between the two subunits, forming a native-like interface similar to the interface between the two beta sheets in the monomer, while another interface is formed by a new long beta sheet, corresponding in sequence to the loop L4 and the D and E beta strands in the B2M monomer. Fibril structure is described as a runaway domain-swapping with inter-molecular disulfide linkage, explaining the formation of odd stoichiometry oligomers. The biological relevance of this mechanism is of course an issue, because of the strong reducing conditions required to open the disulfide bridge into thiols, but similar reactions have been described for other proteins, often through interaction with specific substrates [85].



(a)

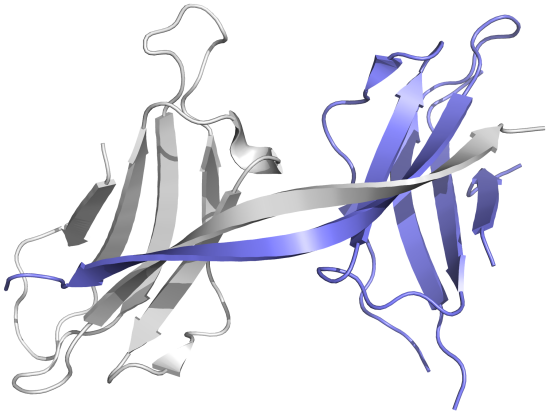


(b)

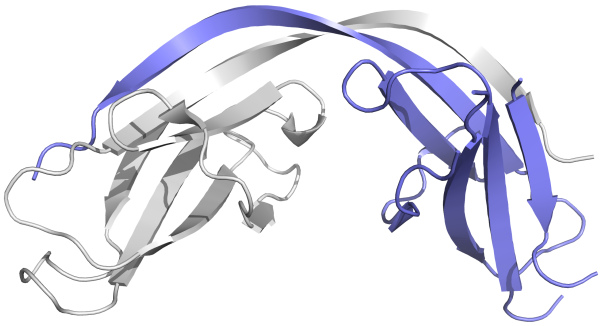
Figure 5.1: Domain swapped dimer of B2M (one monomer in blue, the other in orange) from [108]

In 2011, Domanska and coworkers [49] published the structure of a domain-swapped dimer of B2M, stabilized by nanobodies, shown on figure 5.2. Nanobodies consist of the antigen-binding domain of antibodies, and anti-B2M nanobodies were obtained

from immunization of camels and llamas. Selecting nanobodies which show fibril formation inhibition properties, they were able to isolate and co-crystallize a domain-swapped dimer of B2M, claimed to be an on-pathway intermediate in fibril formation. Contrary to the DTT-induced dimer, this nanobody-trapped dimer conserves its intra-molecular disulfide bridges, making the swapping a much smaller rearrangement, as monomers only swap the C-terminal beta-strand G. In the process, in a similar fashion to that of the DTT-induced dimer, a new inter-molecular beta sheet is formed, through the lengthening of the F beta-strand. As the rearrangement is globally smaller, the hinge is shorter, consisting of only four amino-acids. Even though runaway domain swapping is proposed as a mechanism for the formation of fibrils, the short hinge raises steric problems.



(a)



(b)

Figure 5.2: Domain swapped dimer of B2M (one monomer in blue, the other in gray), stabilized using nanobodies, from [49]

Miranker and coworkers have shown that B2M is a metal binding protein, the binding site identified through site-directed mutagenesis being His31. B2M has a higher affinity to Cu^{2+} than to any other divalent metal ion, and only Cu^{2+} can induce

the formation of amyloid fibrils in physiological conditions [52]. Upon binding with Cu^{2+}, B2M undergoes a conformational rearrangement, with a final state acquired in about one hour at 20°C, a timescale that is not dependent on either copper or protein concentration, suggesting that the limiting factor is the rearrangement itself. This activated state is termed M* by Miranker and coworkers. This M* states is then involved in oligomerization. As observed in the comparison of CD spectra, global secondary structures are conserved for monomers inside the oligomers, implying the formation of oligomers with native-like fold domains. These oligomers are intermediates on the pathway of formation of amyloid fibrils, that can be induced in physiological conditions with concentrations of copper (3µM) that are equivalent to those that can be found in dialysates of patients undergoing long term hemodialysis. If copper-binding triggers the formation of oligomers, and of the subsequent amyloid fibrils, it is not necessary for the stability of the fibrils. While early oligomers disassemble back to monomers upon addition of EDTA, mature fibrils remain stable. Early oligomers, within a timeframe of days to weeks, undergo modifications that makes them insensitive to the addition of EDTA, and can be called "chelate-resistant" [27].

To isolate and crystallize copper induced oligomers, a single point mutation (H13F) is required to reduce aggregation while keeping the copper binding affinity and oligomerization rates. Miranker and coworkers have been able to isolate and crystallize an hexameric form of B2M, induced by copper addition. Interestingly this hexamer has a trans His31-Pro32 peptide bond, while it is in a cis conformation in the wild-type, native state B2M. The isomerization of this prolyl bond may be induced by the binding of copper on His31, and would account for the rate limiting step of the oligomerization process. The hexamer displays a closed-ring structure (fig 5.3), three fold symmetric, demonstrating two classes of interfaces, so that the hexamer can be described as a trimer of dimers. One interface involves interaction between D beta-strands, while the other corresponds to the stacking of ABE beta-sheets, each strand facing its counterpart. The copper ions are not involved in the interfaces, but have binding sites that are strictly intra-molecular, around His31, underlying the fact that it is the conformational rearrangement induced around Pro32 by copper binding that is the key event in the oligomerization of B2M. As a side note, the binding site of the copper around His31 in B2M is very similar to the binding of copper to the octapeptide region in Prion Protein. Once the trans conformation of the H31-P32 peptide bond is sufficiently stabilized through oligomerization, copper is released, making the oligomer a "chelate-resistant" structure. It has been demonstrated that copper-binding induces a 260-fold increase of the backbone isomerization rate of amides [44].

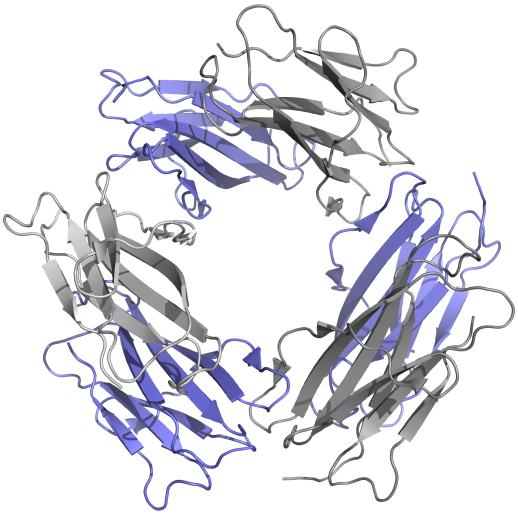
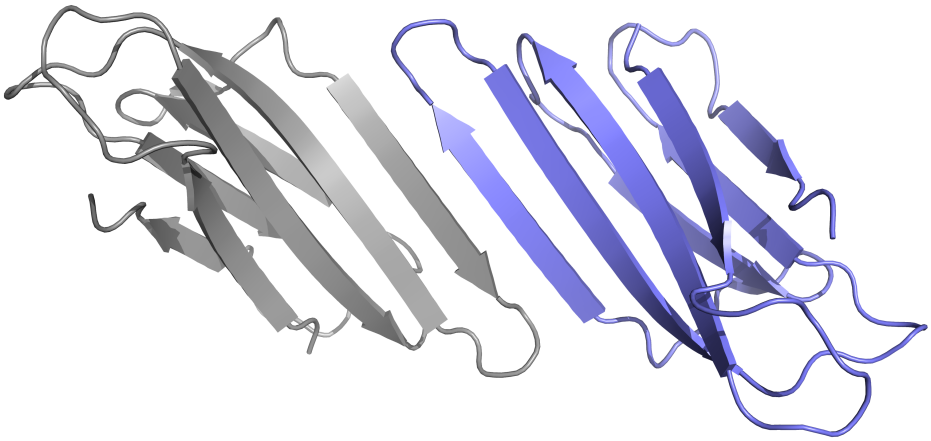
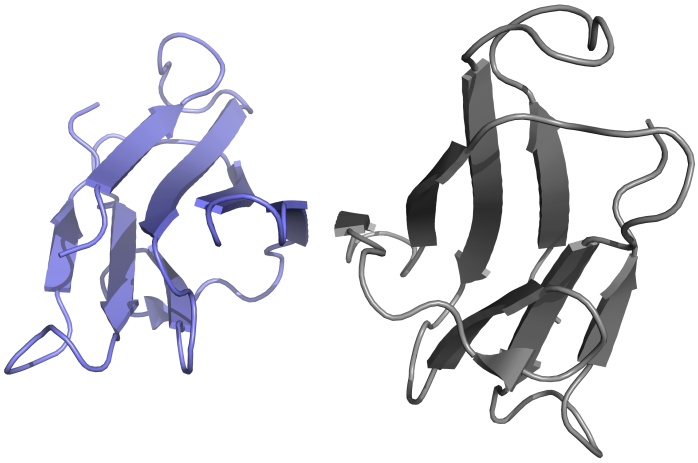


Figure 5.3: Crystal structure of the H13F-B2M hexamer obtained in presence of copper

In the structure of WT B2M, the side chain of His31 is exposed to solvent and in a position orthogonal to the Pro32 amide plane, so that the two residues should be able to form a tetrahedral coordination for copper through the isomerization of the bond. To test this hypothesis, Miranker et al [19] produced a P32A mutant of B2M. Since non prolyl peptide bonds have a minimal probability to be in a cis conformation, P32A should be a good model of a stabilized trans conformation state. And indeed P32A binds copper with a 10000 fold increased affinity compared to WT B2M. In the crystal structure of the P32A mutant, the side chain of His31 is solvent exposed and the His31-Ala32 peptide bond is in a trans conformation. Interestingly, P32A crystallizes in the form of a dimer (figure 5.4). The dimer interface is made of the interaction of the D beta-strands of the two subunits, a D strand that is stabilized with the disappearing of a beta-bulge around Asp53. This kind of interaction is similar to that seen in one of the two interfaces of the hexamer of H13F B2M, and to the beta-sheet hinge of the DTT-induced dimer. In such a configuration, through this D-D interaction, a strong 8-strand beta sheet is formed, with strands ordered as such: ABED-DEBA. This ABED-DEBA sheet forms a hydrophobic stripe, that would be stabilized through the formation of a twisted pair of these sheets, creating the cross-beta core of amyloid fibrils. That kind of arrangement implies an alternating head to head (D-D) and tail to tail (A-A) interaction. Through all these observation, Miranker et al suggest that P32A could be a good in-equilibrium model of the transient M* state, or even of the I_t folding intermediate state (see chapter 12).



(a)



(b)

Figure 5.4: Crystal structure of P32A-B2M dimer obtained in presence of Cu from [19]

5.4 B2M native oligomers

If crystallization is the method of choice to obtain precise structural information, especially on high molecular weight species like B2M oligomers, the crystallization process can be viewed as a selection of a unique structure, while other populated states of the ensemble are discarded. NMR can be of help to assess for the presence of multiple populated states, but it is of course limited in molecular weight. Other methods such as SAXS can characterize higher mass species, but even though signal analysis methods have been developed to deconvolute the signals specific to the different constituents, polydispersity is really an issue. Far-UV CD can be used to detect different species during the denaturation of B2M, but it is not able to characterize them [116].

One technique allows detection and characterization of high molecular weight species in a polydisperse sample: mass spectrometry. It was first used to study B2M in 2004, and showed that different states were populated during the acid unfolding of the protein [24]. The deconvolution of the ESI-MS signals allowed the following of the population of the different states from pH=6 to pH=2. If only the native conformation of B2M can be detected at pH=6, at pH=3,6 where a partially folded state was supposed to be populated [116], a mixture of the native and this partially folded state is detected. At pH=2,6, the two preceding states are still populated, while an acid unfolded state appears.

Based on a difference of the charge state characteristic of those different species, ESI-MS should be even more powerful to detect the formation of oligomers. In 2006, Smith and coworkers [159] followed the formation of B2M fibrils at pH=2,5 and pH=3,6 using ESI-MS. A comparison of the monomer concentration detected using mass spectrometry and the fibril formation using Thioflavin-T fluorescence showed different behaviors at the different pHs. At pH=2,5, the lag phase in the fibril formation correspond to a constant concentration of monomer, while at pH=3,6, during the lag phase, the monomer concentration is already dropping. This is confirmed when looking at the presence of oligomers during the lag phase. At pH=2,5, low populations of dimers, trimers and tetramers are detected, while at pH=3,6, a whole ladder up to more than a decamer is detected during the lag phase. This study shows the ability of B2M to form oligomers of all sizes even at the very low concentration used in ESI-MS (0,02mg/mL ~1,5µM). In a 2010 study, Santambrogio and coworkers [144] showed using ESI-MS that wild type B2M at physiological pH forms dimers, trimers and tetramers at concentration as low as 15µM, which is much lower than the usually observed threshold for non specific aggregation, which is 100µM in ESI-MS. Moreover, the more stable mutant W60G is shown to have a lower oligomerization propensity.

In conclusion, considering either native or specific conditions, B2M has a high propensity to form oligomers. Stochiometries, structure, and stability are diverse, and the

term of oligomeric state ensemble may be more appropriate.

Chapter 6

B2M folding

6.1 A central disulphide bridge

The first factor that was studied in B2M folding was the intramolecular disulfide bridge. Using CD, in presence of 1M Gd-HCl and at pH 7,8, Isenman and coworkers [89] showed that B2M could not reach a native like fold when the disulfide bridge was reduced into thiols, suggesting that the folding of B2M to the native state implied the formation of the disulfide bridge. Yet in 2002, Ohhashi and coworkers [123] using the same technique but without Gd-HCl and at a pH of 8,5 claimed that reduced B2M reached a native-like fold, even though reduced B2M was very aggregation-prone. This shows that the importance of the disulfide bridge is controversial, even though all subsequent B2M folding studies use a protein with the intracellular disulfide bridge formed. Moreover, although the folding of b2-m in vivo occurs in the endoplasmic reticulum and is associated with formation of the disulfide bond, it has been shown that this bond is not broken in the conversion of native b2-m into amyloid fibrils in vivo [11], and almost all models of B2M fibrils have an intact intramolecular disulfide bridge.

6.2 B2M folding studied in real time

The first study in real time of B2M folding was made in 2001 by Chiti and co-workers [37]. The folding was initiated using a 20 fold dilution from a 3M Gd-HCl buffer into a Gd-HCl buffer with a concentration ranging from 0 to 1M, and followed using either intrinsic fluorescence, CD or ANS binding. Three folding phases can be distinguished in the UV intrinsic fluorescence data. First a very fast phase, or burst phase, that happens during the time course of the deadtime of these stopped-flow experimental setups, which is of a few milliseconds, detected due to the difference between the

extrapolation of the intrinsic fluorescence at t=0 and the value corresponding to the unfolded protein. Second is a fast phase, that is completed in less than two seconds at 30°C, and that can be followed by the above mentioned techniques. Finally a slow phase appears, with a time constant of $3.10^{-3}s^{-1}$. The corresponding model is shown on figure 6.1.

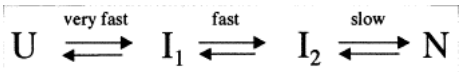


Figure 6.1: Sequential B2M folding model from [37]

In the CD data, the burst phase and the slow phase are detected, with similar timescales as those observed by fluorescence, but the fast phase could not be observed. The far-UV spectrum obtained after the burst phase is significantly different from the native state spectrum. There is a lack of signal in the band close to 200 nm that suggests the presence, at this stage of the folding, of unstructured regions that have to undergo a conformational rearrangement to achieve a native-like content of beta-sheet structure. The absence of fast phase in this data suggests that I_1 and I_2 are similar in terms of secondary structures, with regard to far-UV CD. In presence of ANS, all three phases are observed, with similar time scales. While the ANS-fluorescence of the native and unfolded states of B2M are similar, there is a significant increase during the burst phase, followed by a two-phase decrease. This increase of ANS-fluorescence is characteristic of the exposition of hydrophobic residues on the surface of the protein[1]. In conclusion, the folding mechanism according to this article consists of a burst-phase collapse, during which most secondary structures are formed, yet followed by a two-phase reduction of unstructured regions of the protein. The model proposed by the authors is therefore sequential and shown in figure 6.1.

Another article describes similar intrinsic fluorescence data, but comes to quite different conclusions on the folding mechanism of B2M. In 2006, Jahn and coworkers [92] have studied the folding of B2M, initiated by dilution, with a final concentration in Gd-HCl ranging from 0,2 to 1,6M. Two phases appear during the folding, with timescales similar to those observed by Chiti et al. A global fit shows that the simplest model to correctly fit the data requires five states, as shown in figure 6.2. From the kinetic time constants, the burst phase observed by Chiti clearly appears as the transition from I_c to N, even if the authors did not search for and measure a burst phase. The fast phase folding would correspond to the U_t to I_t, I_c to I_t, and I_c to N transitions, since U_t and I_c should be the only species significantly populated after the burst phase.

[1]The interaction between ANS and the folding intermediate state can actually be mapped using real-time NMR techniques, see figure 14.6 in annex.

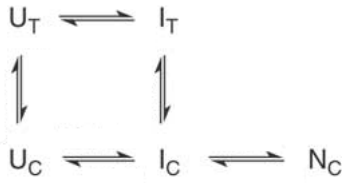


Figure 6.2: Parallel five-state model from [92]

The Jahn and coworkers' parallel model is therefore based on a statistical analysis, while Chiti and coworkers use three qualitative analyses to support their sequential model. First is the fact that, in their four state model, a complete folding (U to N) during the fast phase should imply that a fast-phase unfolding, similar to the fast phase folding, is observed when Gd-HCl concentration is within the transition between folded and unfolded B2M. This is not incompatible with the Jahn et al. model, as the fast phase corresponds to different transitions in which only one leads to the formation of N. A second argument is the fact that the near-UV CD spectrum obtained just after the fast phase is very similar to that of the unfolded protein in a certain region. In the Jahn et al. model, after the fast phase only I_t and N are populated, the population of N being roughly half of the original population of U_c. Due to the noise in the near-UV spectrum, a small population of N may not be detectable, and would make the Chiti et al. compatible with the Jahn model, and this will be corroborated by the NMR data, as described later on. A third argument is that when a double jump experiment, consisting of an unfolding performed just after the fast phase, no slow phase unfolding is observed. This slow phase unfolding should arise from the unfolding of the protein that would be in the N state after this fast-phase. But again, if N is lowly populated, this slow phase may not be detectable. In conclusion, the Chiti et al. data and the Jahn et al. model are not incompatible, and the latter using more data than what was available to Chiti and coworkers, it should be considered as a more precise description of the folding mechanism of B2M.

These observations were complemented by Sakata and coworkers in 2008 [143], using again intrinsic fluorescence and extensive double jump studies in supplement to the reproduction of the folding and unfolding studies already performed on B2M. Double jump experiments correspond to either $N \rightarrow U_c \rightarrow N$ or to $U_t \rightarrow I_t \rightarrow U_t$ transitions, or even $N \rightarrow U_c \rightarrow U_t \rightarrow N$, using for this last jump a variable incubation time in the unfolded state. These double jump experiments allow the direct measurement of the time constants of certain processes that are usually coupled to other processes or inaccessible, such as the $I_t \rightarrow U_t$ transition, as the usual unfolding transition is $N \rightarrow U_c$, in which I_t is not involved. Based on this extensive data, that both reproduces and complements Chiti et al. and Jahn et al. data, Sakata et al. propose a four state minimal model (fig 6.3), and discard the I_c intermediate state of the

Jahn model as a statistical artifact. Moreover, Saketa et al. insist on the fact that microscopic rate constants described in the statistical Jahn model should induce a detectable burst phase in the $N \rightarrow U_c \rightarrow N$ double jump experiments, but that such a phase is not detected in any of the data of the three articles.

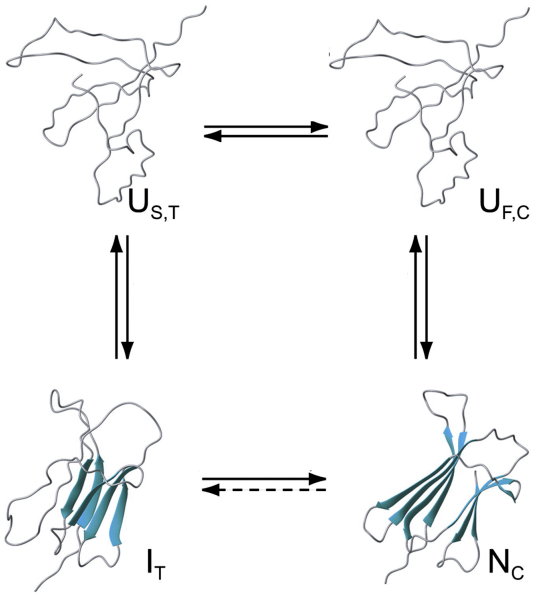


Figure 6.3: Parallel four state folding model from [143]

These studies show how precise and powerful fluorescence spectrometry can be to decypher the folding mechanism of proteins. Yet the main limitation of the technique is the necessary low concentration of protein that must be used, in these studies around 10µM, that does not allow the study of folding in oligomerization-favorable conditions.

6.3 B2M folding followed using NMR

In 2005, Kameda and coworkers [97] studied the folding of B2M using real-time NMR. Yet, using standard HSQC pulse sequences, the temperature had to be brought down to 2.8°C, to cope with the low time resolution of the method. A pH jump was used to initiate the folding, and 1.5M urea was added to the buffer to prevent aggregation of B2M, as the protein remains in a native fold at this denaturant concentration. The dead time in this experiment was 6 minutes, and measurement of one 2D spectrum required 20 minutes, and the whole folding reaction was completed in 15 hours. The first spectrum after the pH jump is completely different from the unfolded state spec-

trum, implying that the unfolded state transforms into the intermediate state within the dead time of the experiment. In this first spectrum, peaks that are specific of an intermediate state can be detected, as they are clearly isolated from the native state peaks. The intensity of these peaks moreover decreases over time, contrary to the native state peaks. Rate constants of intensity growth were more or less the same for all the residues, while the burst phase intensity, ie the intensity ratio between the extrapolated initial value and the final value, was residue dependent. From these burst-phase intensity values, Kameda and coworkers deduce that the structure of the intermediate is close to native in the region that is opposite to P32, and more disorganized close to this residue, which is consistent with the isomerization of proline 32 being the major folding event. Although a two fold lower time constant was observed in NMR compared to CD, it was considered consistent, due to the uncertainty of the measures and the 2°C difference in the experimental temperatures.

6.4 Folding and amyloid formation/aggregation

As was seen earlier, the isomerization of the His31-Pro32 peptide bond is a key element both in the oligomerization process that leads to the formation of amyloid fibrils and in the folding process of B2M. If no structures of the folding intermediate state of B2M exist, due to its short-transient nature, the importance of this isomerization in both events poses the question of the similarities and possible identities in the species involved. Since I_t and M* (aggregation intermediate state, as seen in section 5.3) both have a trans His31-Pro32 peptide bond, could they be the same structural state? Can P32A and DN6, having the same conformation at this position, be used as in-equilibrium models for both intermediate states? Small proteins such as B2M are thought to fold in a rough funnel-like energy landscape [176]. But as the energy landscape represents the ensemble of all the accessible conformations for the protein, the folding funnel is only a part of a larger picture, especially when considering possible formation of oligomers, or interaction with other proteins and constituents of the intracellular medium. Competition between the two phenomenons, the intra-molecular evolution that is folding, and the inter-molecular interactions, leads to an even more complex energy landscape. If folding funnels of small, mono-domain proteins are now quite well-known, our understanding of the energy landscape of protein aggregation is minimal. The formation of weakly-bound oligomers could lead to the formation of an ensemble of conformations of similar energy, equivalent to shallow energy wells, while the formation of highly stable amyloid fibrils would correspond to much deeper wells, possibly even as deep as the one of the monomeric native state. But the picture is even more difficult as even in similar environments, a multitude of fibril morphologies are observed [95]. A representation of such a complex landscape can be found in figure 6.4, extracted from reference [93].

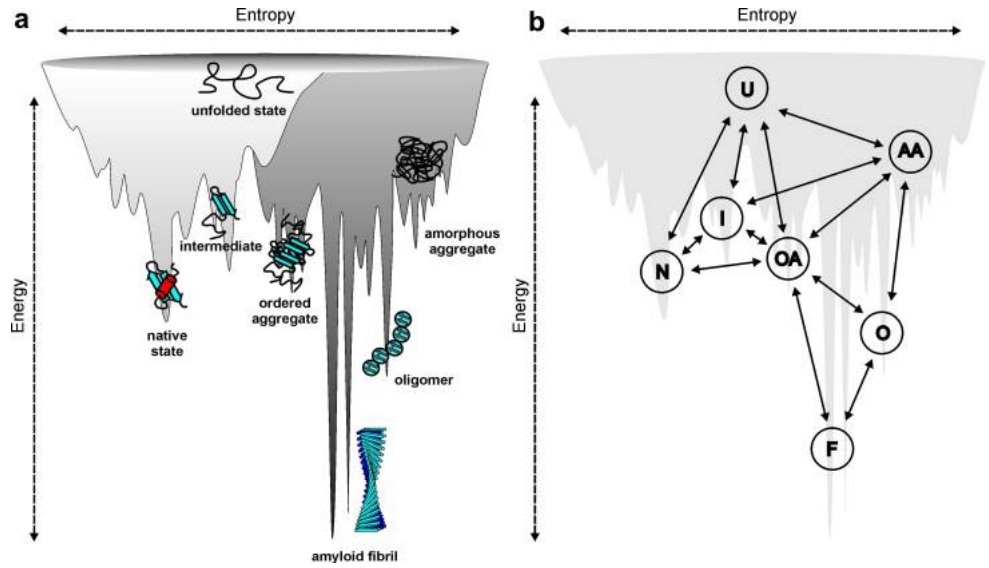


Figure 6.4: (a) Energy landscape of protein folding (in light grey) and aggregation (in dark grey). (b) protein folding and/or aggregation pathways[93].

The concentration of the folding intermediate I_t, enhanced in equilibrium using acetonitrile or during folding, correlates with a much higher fibril formation rate [36]. This fact is consistent with the isomerization of the prolyl bond at P32 being the highest energy barrier. Furthermore, using 2D NMR, DN6 has been shown to have a $^1H-^{15}N$ spectrum similar to that of I_t [55]. Yet P32A, which has a trans conformation and forms oligomers, doesn't show any improvement in the formation of fibril, contrary to DN6 [58]. Whether I_t is inter-converting easily with M*, or that they are exactly the same state is not the most important question for the global understanding of the energy landscape, and a rugged energy landscape can be expected in the region that links the folding funnel and the aggregation.

Part III

The folding of B2M: from kinetics to a folding model

Up to now, most studies of the folding of B2M have used traditional biophysical methods such as UV-fluorescence spectroscopy or circular dichroism. While these techniques provide very accurate kinetic data, they can not offer structural information, and only a limited distinction between different populated states. Folding models have been developed, that derive from the indirect, kinetic detection of intermediate states.

Multidimensional Nuclear Magnetic Resonance (NMR) is a technique that allows a direct characterization of protein states, yet traditional uses of NMR lack the temporal resolution necessary to the study of kinetic phenomena such as protein folding in real time. One solution to this problem is changing some of the thermodynamic parameters of the phenomenon, yet at the cost of physiological relevance, for example through the decrease of temperature, but as shown in section 6.3, even at 2.8°C, traditional NMR methods are not precise enough.

Recent development in NMR methods, along with the availability of high field spectrometers and cryogenically cooled probes providing the required sensitivity, strongly reducing the time needed to acquire correlation spectra, provide another solution to follow the folding of small proteins. Two main drawbacks of real time NMR are the high concentration of protein required, implying large amounts of isotope-labeled samples. The description of these methods and their use for the study of B2M folding is the object of this part.

Chapter 7

Real time NMR: theory and methods

7.1 NMR theory crash course

The NMR theory reminder presented here will focus on the topics necessary for the experimental work described in this manuscript. For further readings, "NMR theory explained" textbook by Jacobsen should be considered for a step by step progressive approach, and "protein NMR spectroscopy" textbook by Cavanagh et al. for a more rigorous and in-depth demonstration.

7.1.1 Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) is a technique based on the interaction of an electromagnetic radiation and nuclei placed in a constant magnetic field. More specifically, the interaction of the radiation and a quantum mechanical property of nuclei, the spin. The spin is an intrinsic property of a particle, as is mass or electric charge. If the spin of a nuclei has a non-zero value S, it may be used as a probe for NMR experiments. Such non-zero spin nuclei that are useful for biomolecular NMR spectroscopy are ^{1}H, ^{2}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P, among many others. A net spin (non-zero) induces a magnetic moment, which can be seen as a small magnet, that will align with constant magnetic fields, and can be perturbed by electromagnetic radiations.

Since magnetic moment is quantized, its accessible energy levels are limited, the number (N) of which is linked to the value of the spin by the relation N=2S+1. For example, for nuclei with S=1/2, like^{1}H, ^{13}C or ^{15}N, only two states are accessible for the magnetic moment. These states are degenerate in absence of an exterior magnetic field. Yet when a constant magnetic field is applied, and due to the interaction of

that field and the magnetic moment of the nucleus, the states no longer have the same energy. A transition from the lower energy level to the higher energy level is therefore possible through the absorption of an electromagnetic radiation with the corresponding energy. This energy is linked to a property of the nucleus, the gyromagnetic ratio, and to the intensity of the external field B0, by relation (7.1).

$$\triangle E = \gamma \hbar B_0 \tag{7.1}$$

To illustrate the interaction of the spin magnetic moment and an electromagnetic radiation, an analogy with classical mechanics can be made, considering the magnetic moment as a magnet. In equilibrium, the magnet is aligned with the external magnetic field B0 along the z axis. Disturbed by the electromagnetic radiation, the magnet will deviate from the z axis. It will then precess around the z axis, at a frequency called Larmor frequency, linked to the gyromagnetic ratio and the intensity of the external magnetic field. This frequency is the resonance frequency that is specific to each type of nucleus. This frequency is modulated by the chemical environment of the nucleus in a process called chemical shift, which provides the possibility to use them as probes of the local environment in a molecule. The motion of electrons in the magnetic field B0 generates a secondary magnetic field. For a given spin, this secondary magnetic field will enhance or oppose the B0 field, and will influence the resonance frequency of that spin, that is therefore shifted from the Larmor frequency. This shift is called chemical shift.

7.1.2 Notions of spin relaxation

Relaxation is the process in which a spin system returns, from an off-equilibrium state, to its equilibrium population distribution. Many mechanisms are possible pathways for relaxation, but all involve interaction of spins with their environment. There are two main mechanisms that influence spin relaxation in solution: dipole dipole interaction and chemical shift anisotropy. Other relaxation mechanism include quadrupolar interaction and scalar relaxation, but won't be developed here.

Dipole-dipole interaction Any nucleus that possesses a non zero spin generates a magnetic dipolar field. Inside a molecule, these individual magnetic fields influence surrounding spins, creating an effective field B_{eff} that is slightly different from B_0. As the molecule tumbles in solution, this field fluctuates, thus creating a mechanism of relaxation. The strength of the dipole-dipole interaction depends on the gyromagnetic ratio, and is therefore more important for proton spins.

Chemical shift anisotropy The local magnetic fields induced by the electronic environment that give rise to chemical shifts are by essence anisotropic. When the

molecule reorients as a result of molecular motion (or tumbling), the local magnetic fields vary. This variation is another mechanism of relaxation.

7.1.2.1 Macroscopic representation of relaxation

The approach introduced by Bloch was conceived to assess for the observations made in early NMR experiments. Therefore it consists of a macroscopic, classical, description of bulk magnetization in a sample. For example, after a 90° pulse in NMR, the bulk magnetization is aligned in the xy plane, the plane whose normal is the direction of the external magnetic field B0, and is called a transverse magnetization. This transverse magnetization precesses around the external magnetic field, creating a current in the receiver coil, producing the signal that is recorded in NMR. It is observed that in accordance with thermodynamic theory, this magnetization "relaxes" to equilibrium in a process that can be decomposed into two mechanisms. The NMR signal is a decaying sinusoidal function called FID (free induction decay). This decay, that corresponds to the decay of magnetization's projection in the xy plane, is called transverse relaxation, and has a characteristic time called T_2. The magnetization recovery along the z axis, called longitudinal relaxation, has a characteristic time T_1, which is longer or at least equal to the characteristic time T_2 of transverse relaxation. T_1 and T_2 are important values for the NMR spectroscopist, as T_1 provides an evaluation of the recovery delay necessary between two NMR experiments, and T_2 is directly related to the linewidth of peaks in NMR spectra.

This Bloch model can assess for most of the simple NMR techniques, and allows to grasp a basic understanding of NMR principles without using quantum theory and the description of spin systems. Yet more advanced NMR experiments such as HMQC cannot be explained using the Bloch vectorial representation of spins, and a more precise description of the evolution of spin systems is necessary.

7.1.2.2 Microscopic representation of relaxation

Rigorous description of NMR relies on quantum theory, using the product operator formalism or the density matrix formalism and the use of hamiltonians. Most common microscopic representation of relaxation is semi-classical: the evolution of the spin system is modeled using quantum theory, while the evolution of the lattice (molecule) is modeled classically. Such a semi-classical representation of relaxation was introduced by Bloch, Wangsnes and Redfield (BWR)[20, 174, 133, 134] between 1953 and 1965.

In a few words, and with the simple example of a coupled two spin system, the quantum mechanics treatment of the spin system leads to four energy eigenstates as displayed on figure 7.1. Transitions are possible on this diagram between eigenstates

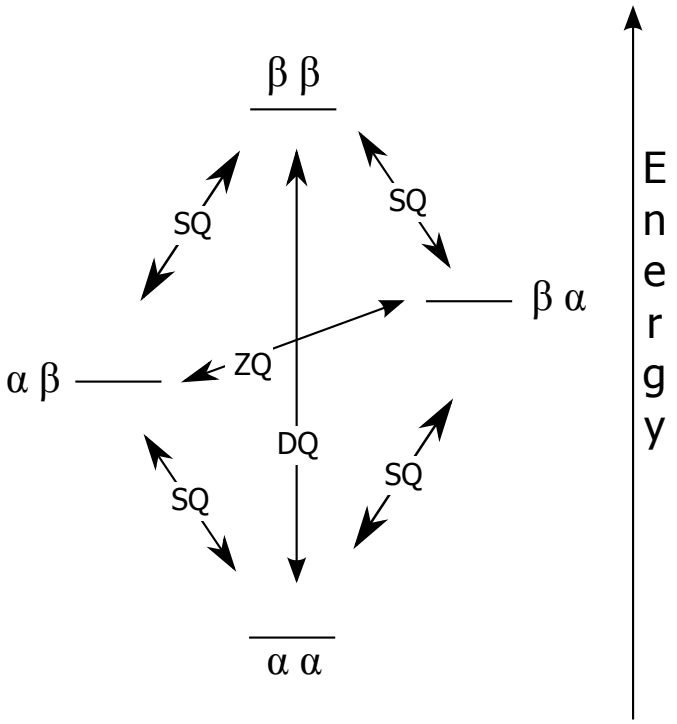


Figure 7.1: Energy diagram of a heteronuclear two spin system, with each spin system having two energy eigenstates $\alpha$and β, leading to the four eigenstates $\alpha\alpha, \alpha\beta, \beta\alpha, \beta\beta$. Single quantum transitions are marked "SQ", double quantum transitions "DQ", and zero quantum transitions "ZQ". DQ and ZQ transitions are often referred to with the generic term multiple quantum transition.

that involve the change in the state of one of the two spins (single quantum transition), or the two spins at once (multiple quantum transition). The transition energy for single quantum transition corresponds to the Larmor frequency of the considered spin, while multiple quantum transitions are linked to the sum and the difference of the Larmor frequencies of the two spins. Transitions between these eigenstates require energy transfer from the environment, that constitutes a local magnetic field. The energy transfer has to be of the specific amount corresponding to the energy of the transitions, as a consequence of quantum theory. This means that the fluctuations of the local magnetic field have to happen at specific frequencies to have an impact on the spin system.

The fluctuations of the local magnetic field can be described by a spectral density function, that corresponds to the energy present within the environment at any given frequency. Since the main cause of local magnetic field fluctuation is overall and internal motion of the molecule, relaxation of the spin system is linked either to the dynamics of intramolecular motions or to the global molecular tumbling of the molecule. The study of the relaxation properties of the protein provides insights on the local dynamics at stake inside a molecule or on the global dynamic behavior of

the molecule.

7.1.3 Relaxation NMR, molecular tumbling, linewidth, chemical exchange

In this thesis, the study of relaxation in a protein is limited to the measurement of effective local relaxation rate constants $R_1 = 1/T_1$ and$R_2 = 1/T_2$ for each amide group of the protein. This gives information on the local dynamics along the amide backbone of the protein. If the R_1 and R2 values are fairly constant along the backbone, this means that internal motions are similar throughout the protein. An estimation of the molecular tumbling correlation time τ_c is obtained through equation (7.2).

$$\tau_c \approx \frac{1}{4\pi\nu_N}\sqrt{6\frac{\overline{R_2}}{\overline{R_1}} - 7} \tag{7.2}$$

Variations in R_1 and R_2 along the backbone may reveal differences in the internal motions, such as more flexible or more rigid parts of the protein. Since R_2 is directly linked to the linewidth of peaks, larger R_2 values may lead to a disappearing of the concerned peaks, for example in presence of chemical exchange, as explained below. Parts of the protein experiencing chemical exchange may completely disappear from the spectra, giving rise to undetectably large peaks.

7.1.4 Relaxation NMR: R_1 and R_2 measurements

7.1.4.1 Inversion-recovery method to measure R_1

R_1 is the relaxation rate of the magnetization along the z-axis. The usual method to measure R_1 is the inversion recovery scheme. First a π-pulse inverts the magnetization along the z-axis. Then a τ delay allows for a partial recovery of the magnetization. Finally, a $\pi/2$-pulse converts the z-magnetization to measurable x-y magnetization (figure 7.2). The intensity of the obtained peaks can be fitted using an exponential function: $I(\tau) = I_{inf}(1 - 2\exp(-R_1\tau))$, which gives the relaxation rate R_1.

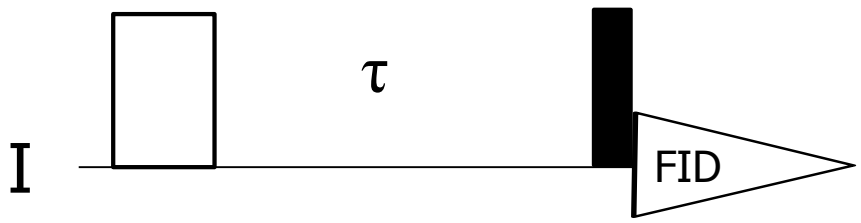


Figure 7.2: Inversion recovery 1D pulse sequence. Filled and open symbols correspond to 90° and 180° rf pulses, respectively.

7.1.4.2 CPMG Spin echo method to measure R_2

In an ideal world, R_2 is directly related to the linewidth of peaks in a spectrum through the decay of the FID. But in the real world, magnetic field inhomogeneity makes the FID decay faster, with a relaxation rate $R_2^* = R_2 + R_2^i$. Therefore the measurement of R_2 is not straightforward. A spin echo sequence can be used to refocus chemical shift evolution that is due to the magnetic field inhomogeneity, as shown on figure 7.3. The π-pulse reverses the chemical shift evolution so that at the end of the second $\tau/2$ delay, the coherence is refocused. Yet not all of the coherence is recovered, due to the intrinsic relaxation rate R_2. Therefore, variation of the τ delay and fit of the peak intensities with an exponential function $I(\tau) = I(0)\exp(-R_2\tau)$ leads to the measurement of R_2. Since diffusion may induce a bias in the measurement, especially for smaller molecules or less viscous solvents, instead of using longer and longer τ delays, the spin echo block can be repeated with a fixed short τ delay. This method is called Carr-Purcell-Meiboom-Gill or CPMG.

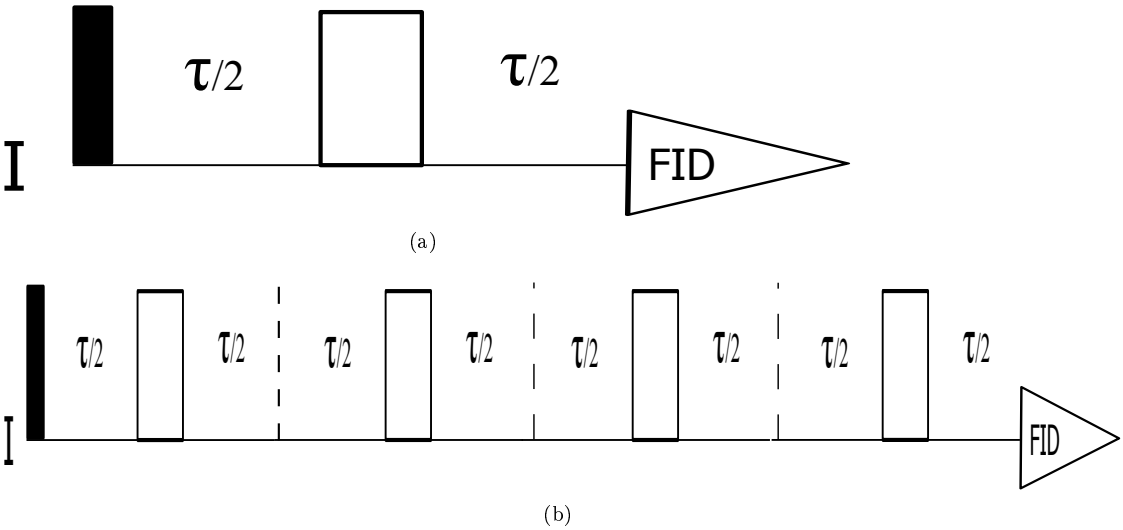


Figure 7.3: Spin echo (a) and CPMG (b) 1D pulse sequence. The τ delay is variable in the spin echo pulse sequence, while in the CPMG pulse sequence, the variable is the number of repetitions of the spin echo. Filled and open symbols correspond to 90° and 180° rf pulses, respectively.

7.1.5 Chemical exchange and relaxation

Chemical exchange is defined as the exchange of a nucleus or nuclei between different environments, either through a chemical reaction or a conformational rearrangement. This includes intramolecular motions, post-translational modifications, oligomerization, exchange of amide protons in aqueous solutions, etc.

For a given nucleus that is in exchange between two magnetically distinct sites, the resonance frequency will be different depending on the site it is in. The key parameters of this exchange are the difference in the resonance frequencies $\Delta\nu$, and the characteristic time the nucleus populates one or the other site $1/k$. The relative positions of k and $\Delta\nu$ on the frequency axis give rise to different regimes.

Slow exchange ($k \ll \Delta\nu$) means that on average each nucleus is detected in either of the two sites, but does not undergo exchange during the "NMR detection". The two sites give rise to dinstinct peaks in the spectrum, at their specific chemical shifts.

Fast exchange ($k \gg \Delta\nu$), on the other hand, means that each nucleus undergoes many exchanges during the NMR detection. Therefore only the average signal is detected in the NMR spectrum, leading to a single peak with chemical shifts that are a weighted average of the chemical shifts of the two sites.

Between those two extreme regimes, a whole range of intermediate exchange regimes are available, depending on how close $\Delta\nu$ and k are. These intermediate exchange regimes are characterized by a broadening of the peaks due to the chemical exchange process. The chemical exchange can be considered as a peak in the spectral density function at the given $1/k$ frequency in the case of a conformational rearrangement.

7.2 Longitudinal relaxation optimized NMR

As seen in section 3.7, 1D ^1H spectra of proteins show strong overlaps between peaks corresponding to different protons, limiting the information that can be extracted from the data. Multidimensional NMR is an elegant way to reduce the peak density of spectra, through resonance frequency correlation. Yet the drawback of the method is that obtaining multidimensional spectra takes time. As seen in section 3.7, multidimensional NMR requires the sampling of indirect dimensions, and therefore the acquisition of numerous scans which grow exponentially with the number of dimensions. For example, typical minimum experimental times are 2 seconds for a 1D spectrum, 4 minutes for a 2D spectrum and 8 hours for a 3D spectrum. The experimental time of a nD experiment T_{nD} is given by equation 7.3 where τ_{rep} is the repetition delay between two scans, N_s is the number of repetitions used for averaging, and N_i the number of increments in the considered indirect dimension. Such a long experimental time is a real drawback to study kinetic processes such as folding, or transiently populated states such as intermediates.

$$T_{nD} = 2^{n-1}\tau_{rep}N_s\prod_{i=1}^{n-1}N_i \tag{7.3}$$

There are two ways to reduce the time requirement of multidimensional NMR: reducing the number of scans, and/or acquiring them faster. Scan number reducing

techniques include spectral aliasing, linear prediction, projection NMR, single scan NMR, or sparse, non uniform data sampling. These techniques won't be developed here, only projection NMR will be shortly presented in section 7.3.

The duration of the acquisition of one scan consists of the pulse sequence duration, the acquisition time, and the recovery delay between scans. This recovery delay is essential to allow the magnetization to relax, since sensitivity is proportional to the available magnetization at the beginning of each scan. Two approaches can be mentioned to account for the maximization of the magnetization within short recovery delays: Ernst angle excitation and longitudinal relaxation enhancement.

7.2.1 Ernst angle excitation

Ernst and coworkers developed[57] a method to maximize sensitivity in a situation in which the recovery of magnetization is incomplete. The application of an excitation pulse with an optimized effective flip angle (Ernst angle:$\beta_{\mathbf{Ernst}}$) can compensate this incomplete recovery. It can be shown that the optimized excitation angle is given by equation (7.4).

$$\cos \beta_{Ernst} = \exp -\frac{T_{rec}}{T_1} \tag{7.4}$$

The optimal excitation angle is 90° when the recovery delay T_{rec} is at least four times the longitudinal relaxation time constantT_1. As shown on figure 7.4, the Ernst angle decreases as T_{rec} is shortened. While at 90° the maximum sensitivity is obtained for $T_{rec} \approx 1,25 \cdot T_1$, the maximum for an excitation angle of 60° is obtained for $T_{rec} \approx 0,6 \cdot T_1$. Moreover, Ernst angle excitation with shorter recovery delays often imply increased sensitivity compared to a 90° pulse with $T_{rec} \approx 1,25 \cdot T_1$. Ernst angle excitation is therefore a method to accelerate the acquisition of NMR spectra, by reducing the recovery delay between successive scans.

7.2.2 Longitudinal relaxation enhancement

The Solomon equations are coupled first order differential equations that describe the time evolution of spin polarization in a coupled spin system. Equation (7.5) shows the Solomon equations for a system of two spins $I_1$and I_2, with I_{nz} the z-component of the angular moment of the spin n and I^0_{nz} its thermal equilibrium value, ρ and σ respectively the auto relaxation and cross relaxation rates.

$$-\frac{d}{dt}\begin{pmatrix} I_{1z} - I^0_{1z} \\ I_{2z} - I^0_{2z} \end{pmatrix} = \begin{pmatrix} \rho & \sigma \\ \sigma & \rho \end{pmatrix} \cdot \begin{pmatrix} I_{1z} - I^0_{1z} \\ I_{2z} - I^0_{2z} \end{pmatrix} \tag{7.5}$$

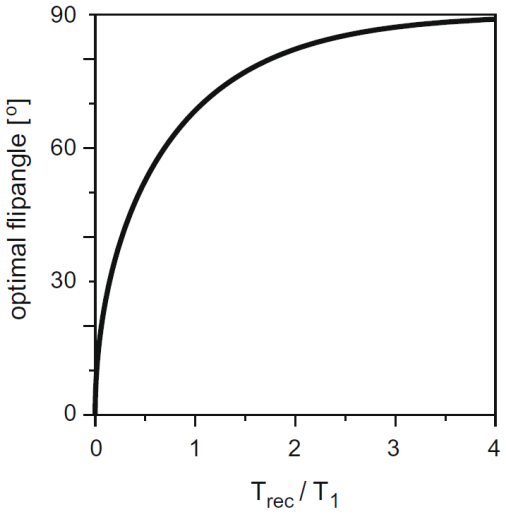


Figure 7.4: Optimized excitation angle (Ernst angle) as a function of the relative interscan delay T_{rec}/T_1

Auto-relaxation describes the dependency of the spin magnetization recovery on the deviation of the spin magnetization to its thermal equilibrium. Cross relaxation describes its dependency on the magnetization state of surrounding spins, through dipolar interaction. For macromolecules such as proteins, it can be shown that cross relaxation can be seen as a diffusion process of the spin polarization within the spin system. If all the spins are in the same state, ie excited unselectively, the diffusion is slowed down to a minimum. But if only a few spins are excited, other spins around are available for this magnetization diffusion process. Therefore selective excitation of spins leads to a dramatic enhancement of the longitudinal relaxation rate.

As a result, selective excitation schemes, in which the pulse sequence will selectively manipulate certain spins of the systems (such as amide or methyl protons), provide a shorter longitudinal relaxation time T_1, and as such the possibility of a faster acquisition of NMR spectra with conserved or even enhanced sensitivity.

7.2.3 SOFAST HMQC experiments

- Heteronuclear multiple quantum coherence (HMQC)

HMQC is an inverse 2D heteronuclear experiment, just as HSQC, inverse because it is ^1H-observed, whereas first 2D correlation experiments were ^{13}C observed. Due to the four fold higher gyromagnetic ratio of ^1H compared to ^{13}C, inverse experiments have an advantage of 8 fold in S/N ratio. HMQC correlates chemical shifts of directly bound atoms H-X, in protein spectroscopy usually amide (H-N) or methyl and aromatic (H-

C) groups. With a simpler pulse sequence than that of HSQC, HMQC relies on zero-quantum coherence (ZQC) and double-quantum coherence (DQC) evolution during t_1.

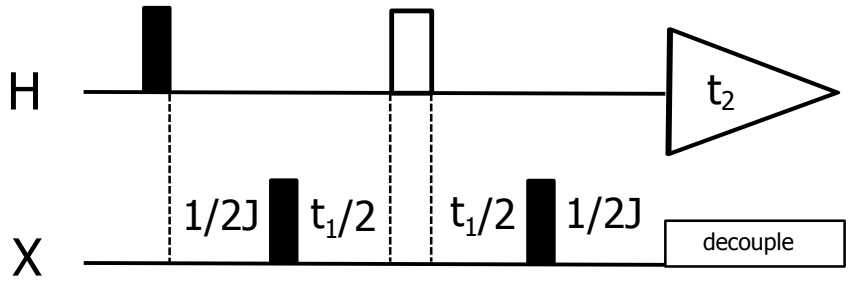


Figure 7.5: Basic H-X HMQC pulse sequence, with π pulses in white and$\pi/2$ pulses in black.

Sequence of a basic HMQC experiment is shown on figure 7.5. First ^1H pulse transforms magnetization from I_z into I_x, and first $1/2J$ delay makes the ^1H magnetization antiphase with respect to S_z: $2I_xS_z$. Chemical shift evolution is neglected, because it will be refocused with the last delay. The $S\pi/2$-pulse leads to a combination of ZQC and DQC: $-2I_xS_y$. Chemical shift evolution of the I spin is refocused during the t_1 delay by the ^1H π-pulse, therefore only S chemical shift occur during the ZQC and DQC. The last $S\pi/2$-pulse converts back the ZQC and DQC back to ^1H antiphase coherence $2I_xS_z$, which is refocused during the final $1/2J$ delay, leading to the detection of both the ^1H and ^{13}C or ^{15}N frequencies.

- SOFAST-HMQC

The band-Selective Optimized Flip-Angle Short-Transient heteronuclear multiple quantum coherence (SOFAST-HMQC) [148] experiment yields significantly increased sensitivity for short acquisition times when compared to other existing techniques. It exploits longitudinal relaxation enhancement and Ernst angle excitation to enhance the repetition rate at which data can be collected.

SOFAST-HMQC is not limited to ^1H - ^{15}N correlation spectroscopy, but can be used to obtain ^1H - ^{13}C correlation spectra of different functional groups of aliphatic or aromatic protons. One example is the specific correlation spectrum of methyl protons, that have a good spectral separation (-0.5 to 1.5 ppm) from other aliphatic or aromatic protons. Methyl protons are valuable probes as they are mostly located inside the hydrophobic core of proteins, that may fold differently from the global backbone probed by the amide protons. Due to their rapid rotation, methyl groups possess three equivalent protons, yielding a theoretical three fold increase in the intensity of the signal, allowing the use of lower protein concentrations or the detection of states that have a lower population. Both ^1H - ^{15}N and ^1H - ^{13}C SOFAST HMQC [148] pulse sequence are presented on figure 7.6.

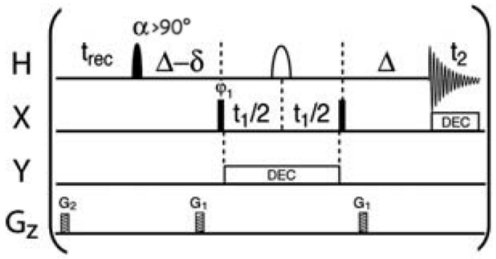


Figure 7.6: SOFAST-HMQC pulse scheme, X and Y can be either ^{15}N or ^{13}C to use amide or methyl protons as probes respectively. Selective pulses are round shaped, π pulses are shown in white, black pulses are $\pi/2$ pulses if not otherwise stated.

7.2.4 SOFAST and FTA-SOFAST

The turbulence induced by the fast mixing of two buffers poses problems when used with pulse schemes containing pulsed field gradients used for selective refocusing, due to sensitivity to translational diffusion effects. These effects can be attenuated using a Fluid Turbulence Adapted pulse sequence, such as FTA-SOFAST. In the standard SOFAST pulse sequence, gradients enclose the indirect evolution time T_1, inducing delay between gradients that can go up to 20-30ms, leading to incomplete refocusing and the concomitant loss of magnetization. To avoid this slow build-up of the signal because of fluid turbulence, the pulsed-field gradients can be moved to the end of the pulse sequence, with an inter-gradient delay fixed to 1 to 2ms, and an additional ^1H refocusing pulse can be added. This sequence (figure 7.7) has proved excellent water suppression abilities and high sensitivity, reducing the dead-time to 1 or 2 seconds, while using the standard SOFAST pulse scheme leads to a dead time of up to 15 seconds [147].

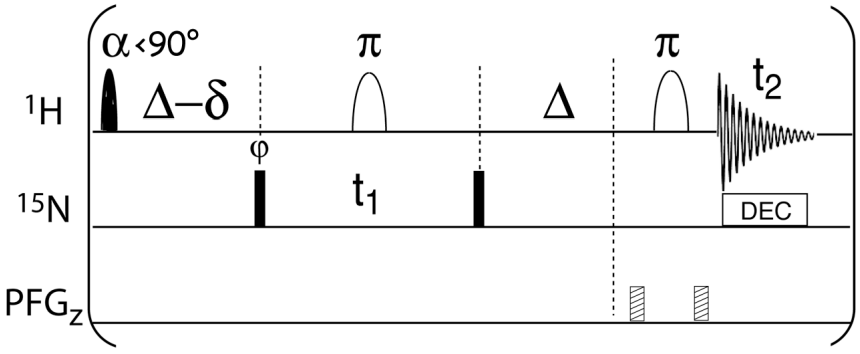


Figure 7.7: Fluid turbulence adapted SOFAST pulse scheme. Selective pulses are round shaped, π pulses are shown in white, black pulses are $\pi/2$ pulses if not otherwise stated.

7.2.5 BEST experiments

Band-selective Excitation Short-Transient (BEST) relies on selective pulses to achieve a longitudinal relaxation enhancement as previously described. This strategy can be applied to a whole set of multidimensional experiments, like the HSQC- and TROSY-type experiments. BEST provides sensitivity enhancement that allows the reduction of the experimental time needed to acquire multidimensional spectra, for example HNCO or HNCA spectra that are used to assign the resonances of a protein. The SOFAST experiment is in this respect a particular case of the BEST strategy, with the added sensitivity enhancement based on the use of Ernst angle excitation.

7.3 Projection NMR

Projection NMR allows the measurement of an experiment of dimension n using a series of spectra of dimension k (with $k \leq n-1$)[102, 163, 155]. Such an experiment is noted $(n,k)-D$ experiment. Projection NMR is based on a particular scheme of spectral sampling of the indirect dimensions, which are sampled along a tilted axis that form projection angles $\alpha_i$with indirect dimensions t_i. In a simple case of a H-X-Y$(3,2)-D$ experiment, and a projection angle α, we have $\alpha_1 = \alpha$ and $\alpha_2 = \pi/2 - \alpha$, figure 7.8 (a). The spectrum obtained is the projection of the 2D X-Y spectrum along an axis that forms the angle α_1 with X and α_2 with Y.

Experimentally this is obtained using the same time variable t_1 for all projected indirect dimensions, with time increments linked to that of the first indirect dimension: $\Delta t_i = \lambda_i \Delta t_1$ with $\lambda_i = \cos\alpha_i / \cos\alpha$. In this simple H-X-Y$(3,2)-D$ experiment, $\lambda = \tan\alpha$ and observed indirect chemical shifts are given by relation (7.6).

$$\begin{cases} w_{+\alpha} = w_X + \lambda w_Y \\ w_{-\alpha} = w_X - \lambda w_Y \end{cases} \tag{7.6}$$

Direct analysis of this data allows the recovery of the Y chemical shift using the central peaks method: an H-X spectrum is recorded, equivalent to a projection with $\alpha = 0$, and for each peak in this H-X spectrum, symmetrical peaks are looked for in the $+\alpha$ and $-\alpha$ spectra. When found, relation (7.6) gives w_Y from λ and $w_{+\alpha}$ or $w_{-\alpha}$, see figure 7.8 (b).

7.4 B2M folding experiments: standard procedures

Lyophilized wild-type B2M is re-suspended in an unfolding buffer (1.5M Urea, pH=2). The folding is initiated by a pH jump induced by the addition of a refolding buffer

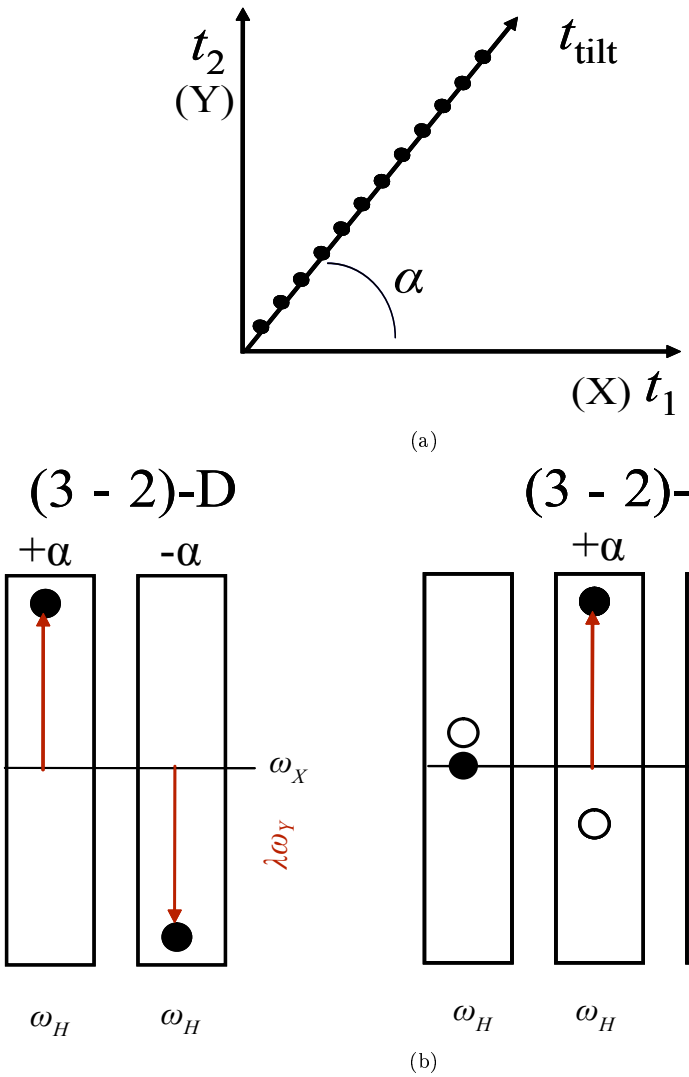


Figure 7.8: (a) H-X-Y$(3,2)-D$ experiment sampling scheme (b) The central peaks method to find the ω_Y by symmetry, using λ and ω_X, for a ^1H isolated peak (left) or a ^1H degenerate peak (right).

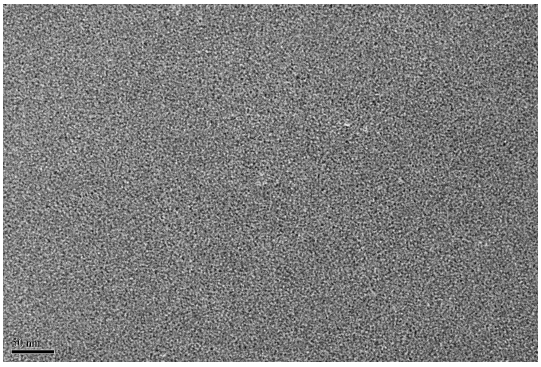
(Phosphate 300mM pH=7.2 or HEPES 900mM, pH=7.2). The measured pH after the pH jump is 6.8. The mixing of the unfolded protein in its buffer and the refolding buffer is made in a 3:1 respective proportion. For example in NMR experiments B2M is re-suspended in 300µL of the Urea buffer, and 100µL of the HEPES buffer is added to initiate folding. The concentration chosen initially for B2M in NMR experiments is 750µM final, and thus an initial unfolded concentration of 1mM. Such a high concentration implies the use of 3.6mg of protein for each folding experiment.

First experiments showed that in these conditions and at this concentration, wild-type (WT) B2M had a strong propensity to aggregate during the pH jump, leading to a sample filled with a milky white precipitate, even in the presence of 1.5M urea. This precipitation was hypothetically due to a seeding by aggregates formed since or already present as soon as the re-suspension step of the process. The elimination of such aggregates or seeds should then eliminate the precipitation. To achieve this, two strategies were developed: filtration and centrifugation. Filtration through 0.2µm filters can be done just before the experiments, and leads to elimination of the precipitation, but at the cost of sample losses both in volume and in concentration. For this reason, preparative ultracentrifugation was used as a standard method for all experiments. The technique consists in the centrifugation of the unfolded sample at 120 000 g for at least 45 minutes (using a Beckman Optima TL ultra-centrifuge), the supernatant being used for the experiment, while the aggregates that are centrifuged to the bottom of the tube are discarded.

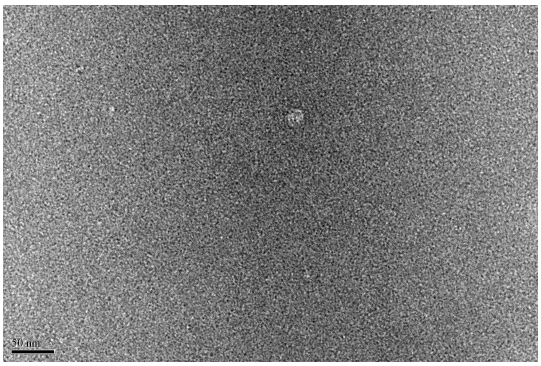
This method allows elimination of all stable aggregates larger than ~25nm in diameter as shown using electron microscopy (figure 7.9), while leading to minimal losses in volume or in protein concentration.

The mixing of the two buffers is performed using an adhoc device, a syringe injector. One main objective in these kinetics studies is to minimize the initial dead time during which data are not collected. When the mixing can be done directly inside the measurement apparatus, the dead-time can be reduced to less than a second, as shown in the description of the fast-mixing syringe injector for NMR experiments. The dead-time can be longer than just the mixing time due to the experimental time needed to obtain the first data point. For example 15 seconds are needed in NMR to obtain the first HMQC spectrum after the pH jump, so even if the mixing is synchronized with the beginning of the acquisition, with a few seconds as a margin of error, dead-time can be considered to be of the order of 10 seconds.

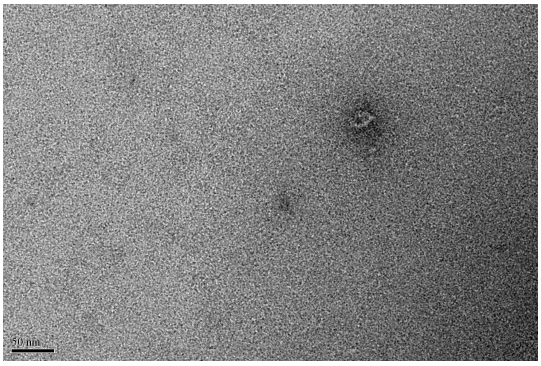
After the pH jump and the dead-time, the kinetic evolution of refolding B2M can be followed through the repetition of acquisitions. NMR being harmless to the sample, acquisitions can be performed at the limiting rate of the technique, every 15 seconds, giving very detailed data on a process that reaches its final state on a timescale of the order of one to six hours depending on the temperature.



(a) Just after pH jump



(b) 15 minutes after pH jump



(c) 3 hours after pH jump

Figure 7.9: Representative electron microscopy measurements that were made at 25°C, using negative staining with sodium silico-tungstate, at different times during the folding, with samples diluted 200 fold, from 750µM, just before measurement. Measurements show the inexistence of aggregates larger than 25nm in diametre along the folding.

7.5 Fast Mixing Device

In order to obtain information on the folding process as soon as possible after its initiation, the initiation step must be as fast as possible. To realize a pH jump, as required here, two solutions have to be mixed and the final sample put into the NMR magnet to start the experiment. If the mixing is done by hand, using a glass Pasteur pipette for example, adding the time needed to fit the NMR tube in the floater, and the time for it to reach the measuring coil's position, the dead-time can be estimated to ~30 seconds to a minute. One way to reduce the dead-time is to perform the mixing directly inside the magnet, using a syringe injector, described on figure 7.10 [147].

This fast-mixing device is composed of a long Teflon tube, that is fixed inside the NMR tube on one side, and goes outside the magnet up to a syringe on the other side. The diameter of the Teflon tube is reduced at its "NMR tube" end, to increase the velocity of the injected liquid and favour a better and faster mixing. This diameter-reduced tube is slightly inserted in the meniscus of the liquid present in the NMR-tube, again to favour a better mixing. The NMR tube used is a Shigemi, its plane bottom being favourable for the formation of turbulence upon the injection, making the mixing even more efficient. Shigemi tubes also have the asset of limiting the volume of sample needed for experiments down to 350-400µL. The Teflon tube is maintained in position inside the Shigemi tube using Teflon mountings, made of a Teflon tube of outer diameter that closely matches the inner diameter of the Shigemi tube. Using such a syringe injector, the mixing can be achieved in less than 100ms, as shown by dye injection, and taking into account the turbulence that appear in the sample upon mixing, with the use of an adapted pulse scheme, experimental dead-time can be reduced to less than two seconds.

7.6 Labeled B2M protein production

One of the assets of B2M, being a good model for amyloid and other misfolding studies, can also become a real drawback. Since it is easy to misfold, or said the other way, difficult to fold correctly, B2M can be difficult to produce for biophysical studies. If some of the most common techniques such as UV-fluo or circular dichroism do not require more than a few milligrams of protein, extensive studies using crystallography or even worse NMR may require up to hundred milligrams of protein. Therefore obtaining a robust recombinant protein production protocol is crucial for these techniques.

NMR of biological samples specifically requires isotopic enrichment of the protein, to introduce ^{13}C and ^{15}N in much higher proportions than in nature, for a higher signal to noise ratio of NMR spectra. ^{13}C natural abundance is only circa. 1%, and ^{15}N is only circa 0.4%, which implies for example for a 2D CH correlation experiment,

about 100 fold less labeled protein than unlabeled protein for the same signal to noise ratio. For the best control of isotopic enrichment, minimal growth media are used for production in E.coli, meaning media that contain the minimal inorganic salts necessary for bacteria growth and thus protein production, in which is added an isotope enriched carbon and/or nitrogen source. The use of such a minimal media usually reduces the yields of protein production, making the optimization of the production protocol even more important.

The major difficulty that was faced during ^{15}N or $^{13}C-^{15}N$ enriched B2M production was that the central disulfide bridge would not form correctly, resulting in a misfolded protein. This could only be detected by NMR and not exclusion chromatography, implying that the misfolding was not disturbing the global fold of the protein. Several oxidizing conditions were tested during the refolding phase of the purification of the protein, two processes were proven effective, either a long folding rest in 6M Urea at low temperature in presence of gazeous dioxygen, or a flash refolding procedure performed at low temperature. The latter appeared much more efficient even though time consuming due to the protein concentration step necessary, and was therefore used with a yield of 50mg per liter of culture for ^{15}N B2M and 20mg per liter of culture for $^{13}C-^{15}N$ B2M. Such a protocol required at least two weeks of work per liter of culture, from bacteria growth to size-exclusion purification. The extensive protocol is shown in section VI of the annex.

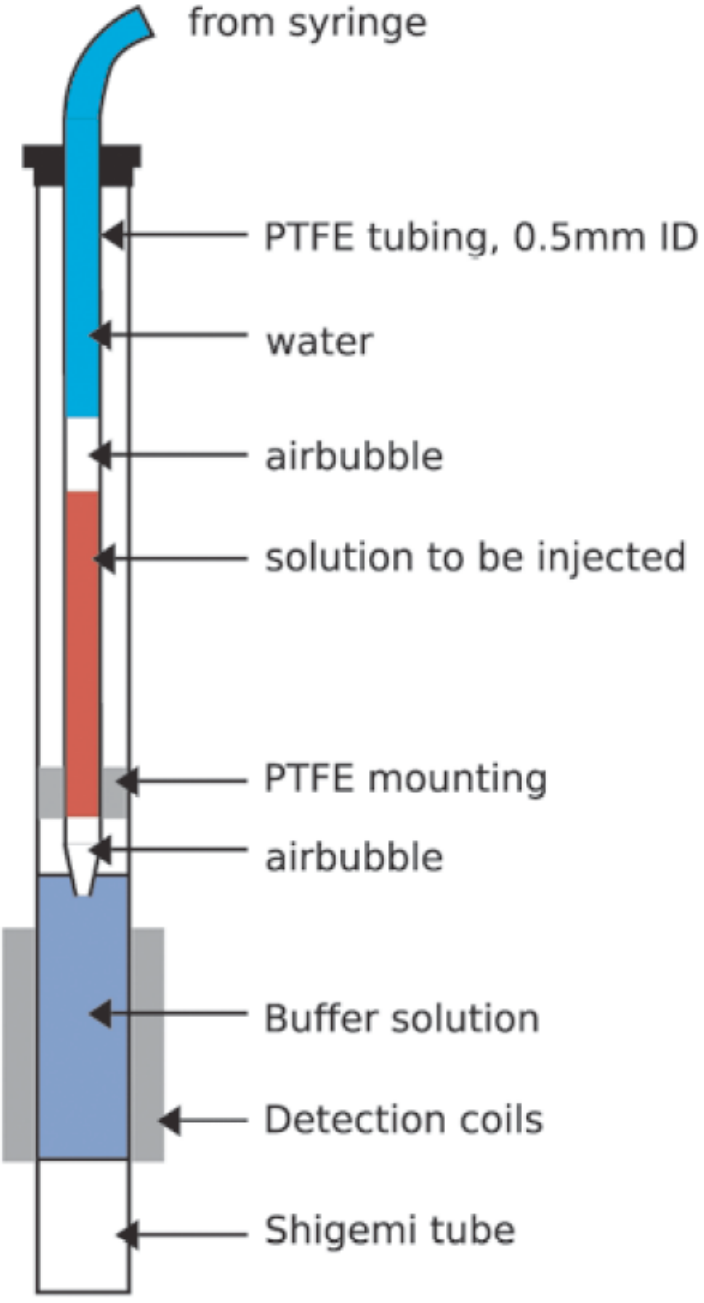


Figure 7.10: Scheme of the fast mixing device used inside the NMR magnet, from [147]

Chapter 8

Folding of B2M followed using real time 2D SOFAST NMR

8.1 Standard experiment:

8.1.1 Data acquisition and processing

NMR experiments were performed on a Varian INOVA spectrometer operating at 800MHz (^1H frequency), equipped with a cryogenic probe. SOFAST pulse scheme allowed to reduce the overall single-scan time down to 100ms. Nitorgen decoupling was achieved using a WURST-40 train [105], while the States-TPPI [113] scheme was employed for sign discrimination with respect to the nitrogen carrier frequency. Each experiment was performed with only one scan per T_1 increment, and phase cycling was performed in subsequent spectra to eliminate artifacts (see 8.1.2). 2D spectra were acquired every 15 seconds. The experiments were processed with NMRPipe. In the indirect dimension F1, the original data set of 60 real points was extended by linear prediction (30points) and zero-filled to 128 points, and a squared sine-bell apodization function was employed, with a shift of $\pi/4$. In the direct dimension F2, a sine-bell function with a shift of $2\pi/5$ was employed. Following two-dimensional Fourier transform, a fourth order polynomial baseline correction was applied to the spectra.

8.1.2 Data analysis

Subsequent spectra were summed to account for the phase cycling. Applying the phase cycling in a two-scan acquisition divides by two the time resolution of the experiment, while phase cycling can be implemented on consecutive single-scan spectra with the

same artifact elimination efficiency. Thus the first step of the data analysis is to sum spectra 1 and 2, 2 and 3, ..., using ad-hoc python scripts and NMRpipe. The only drawback of this method is that the apparent dead time must be considered as half an acquisition longer than that of a standard single scan acquisition. Spectra are then inspected using NMRview.

8.1.3 Results

The following conditions are used for this standard experiment: concentration is 750µM and temperature 25°C, while the folding buffer is the phosphate buffer. While at pH2 B2M has a typical unfolded protein HMQC spectrum (figure 8.1), immediately after the pH jump the scattering of the amide proton resonances implies that B2M adopts a folded conformation. Moreover the unfolded state resonances have completely disappeared even in the first spectrum acquired after the pH jump, showing that the collapse of the unfolded state to a folded state is faster than the time resolution of the experiment (fig 8.1). When comparing the first spectrum after the pH jump and a spectrum of B2M after the folding is completed, a folding intermediate state clearly appears, with resonances that can be either isolated from the native state ones, or with a partial or total overlap.

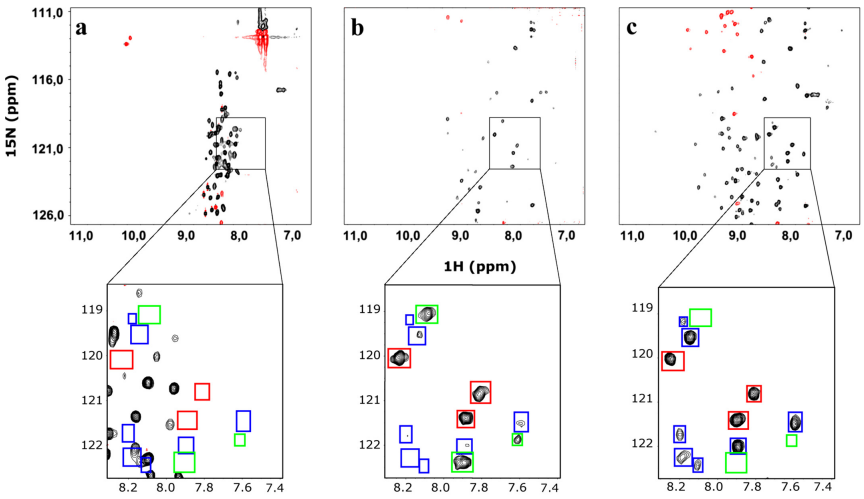


Figure 8.1: Spectral evolution along the folding: (a) Unfolded protein, (b) Just after pH jump, (c) Native state. Peaks highlighted in blue correspond to isolated native peaks, in red to peaks that overlap in the native state and in the intermediate state, n green to isolated intermediate state peaks. The concentration is 750µM and temperature 25°C, while the folding buffer is the phosphate buffer. The spectra were acquired on a 600MHz Varian INOVA spectrometer using the SOFAST pulse scheme, as described in section 8.1.1

Two classes of peaks can be easily drawn: isolated intermediate state peaks (class I) and isolated native state peaks (class N). The build up of the N class signal accounts

for the population of the N state, while the decay of the I class signal accounts for that of the I state. A third class can be distinguished, containing peaks that overlap, fully or partially, in the I and N states, thus the term I+N for this third class. Peaks were classed I+N only when they were isolated from other peaks, and as such could be linked to a single residue resonance. The signal of this I+N class provides useful information on the presence of additional NMR-invisible folding intermediates, or in the difference of local dynamics between I and N states. First result of this folding experiment is that kinetics of signal decay or build-up are similar inside each class (figure 8.2). Therefore, the volume of the peaks of each class was measured in all the 480 spectra using nmrView and TCL scripts, and summed to obtain three signals, one for each class. Results are shown in figure 8.3.

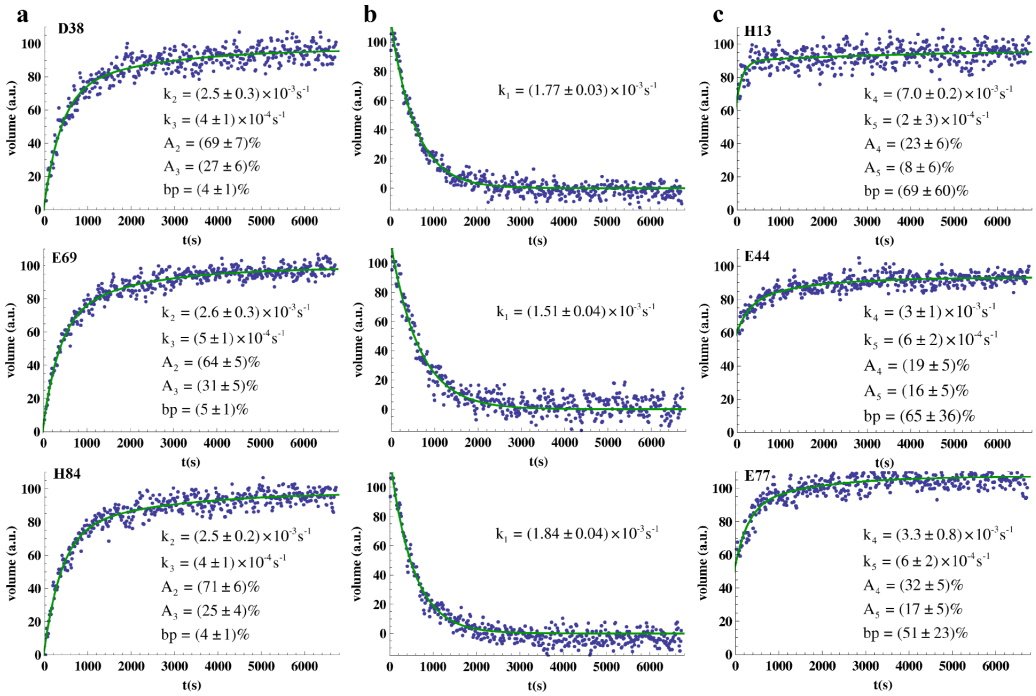


Figure 8.2: Examples of the evolution of different individual peaks within the three peak classes N, I, I+N. The volume is normalized to the final value. Time constants and other fitting parameters and corresponding standard error are shown, as obtained using Monte Carlo simulations.

The native state arises within timescales similar to those observed in UV-Vis spectroscopy. Yet interestingly a bi-exponential fit is necessary to assess for the evolution of the NMR signal, that was not seen in UV-Vis experiments. This implies a folding mechanism more complex than a simple two-state model. On the contrary, the decay of the I class signal can be fitted with a mono-exponential function. Furthermore, the signal of the I+N class is not constant as would be expected in a two-state model, but is rising as a bi-exponential function of time with similar time constants as the N

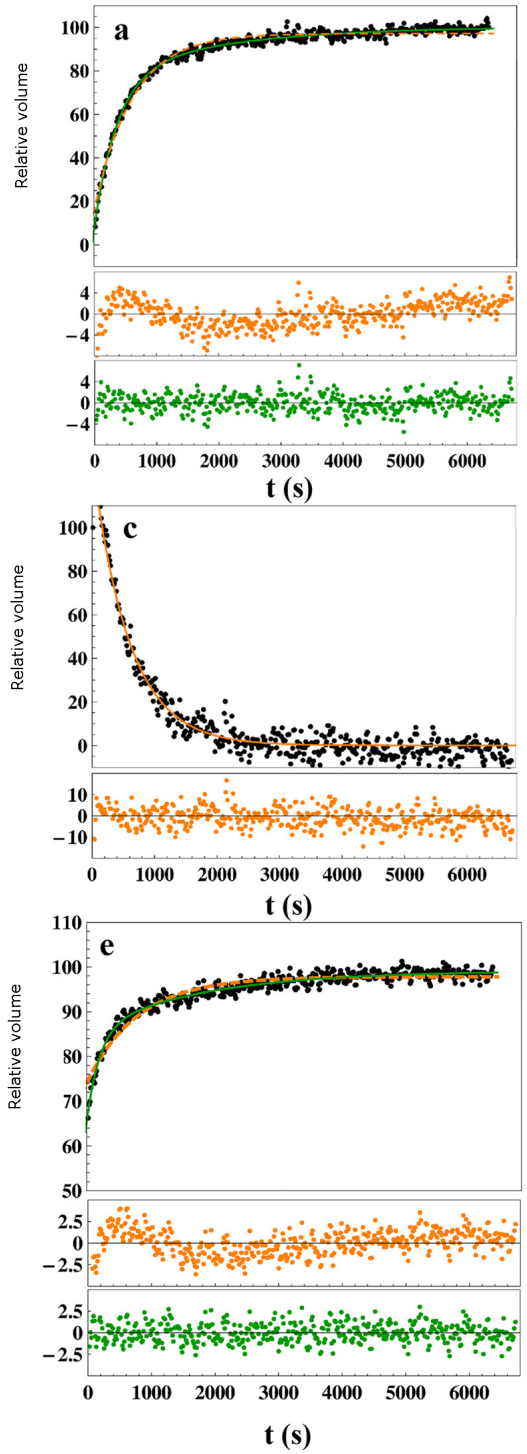


Figure 8.3: Folding of B2M in a Phosphate buffer at 25°C, using peak class N (a), I (c), I+N (e). The volume of the peaks is summed inside each class and is normalized to the final or initial value (black); residuals corresponding to mono-exponential (orange) and bi-exponential (green) fitting functions are shown.

class. The observation of a bi-exponential build-up of the N class signal and the not-constant signal of the I+N class both point toward the existence of another folding intermediate, that is not visible in NMR. Another explanation for the not-constant signal of the I+N class would be differences in the relaxation properties of the I state. Differences in the internal mobility, the dynamics of the I state would result in different spin relaxation properties, and thus in different NMR peak intensities (or volumes) for similar populations of the I and N states.

The evaluation of the line-widths, that are related to the relaxation properties, show no significant differences between I and N state peaks. Furthermore, when performed with a longer recycling delay d1 (thus indirectly monitoring $^1H\ T_1$), the folding spectra display the exact same signal build-up for the I+N class as with the usual shorter d1 (figure 8.4). This rules out the possibility that the missing intensity for the I+N class is due to different relaxation properties of a monomeric intermediate state. A second intermediate state is populated during B2M folding, with a population up to 34% just after the pH jump, a state that is invisible in NMR, either because of the dynamics of the state or because it is an oligomer.

Another fact highlighted by these folding results is the burst phase that exists for the N state: initial population, just after the pH jump and a dead-time of a few seconds, is not close to zero. The N state is populated, as derived from zero-time extrapolation of the N class signal, as much as 7% at the beginning of the folding, just after the pH jump. This burst phase, during which some B2M molecules fold directly and very quickly from the unfolded state to the native state can be linked to the description of a fast folding phase observed in UV-vis [97, 37]. The populations concerned are significantly different, but this can be explained by different experimental conditions, especially different solvents and temperature. Still in the article by Kameda et al. [97], in which they use conventional HSQC pulse sequences to follow the folding at 2.8°C, only a mono-exponential evolution is described for the population of the native state during folding. This can be explained, aside from the different temperature that may dramatically affect the folding mechanism, by the lower time resolution of conventional HSQC with respect to sofast HMQC, that would not allow to go as deep in the description of the fine details of the different states' population evolutions.

	$k_1(s^{-1})$	$k_2(s^{-1})$
N	$(2.5 \pm 0,3) \cdot 10^{-3}$	$(4.7 \pm 0,6) \cdot 10^{-4}$
I	$(1.70 \pm 0,03) \cdot 10^{-3}$	-
I+N	$(4.7 \pm 0,4) \cdot 10^{-3}$	$(5.3 \pm 0,4) \cdot 10^{-4}$

Table 8.1: Kinetic time constants for the folding of B2M in a Phosphate buffer at 25°C, with corresponding standard errors evaluated using Monte Carlo simulations.

A model for the folding mechanism can be derived from the kinetic time constants summarized in table 8.1. With three independent time constants (two for N, that are similar to those of I+N, and one for I), two-state models and models with two folding

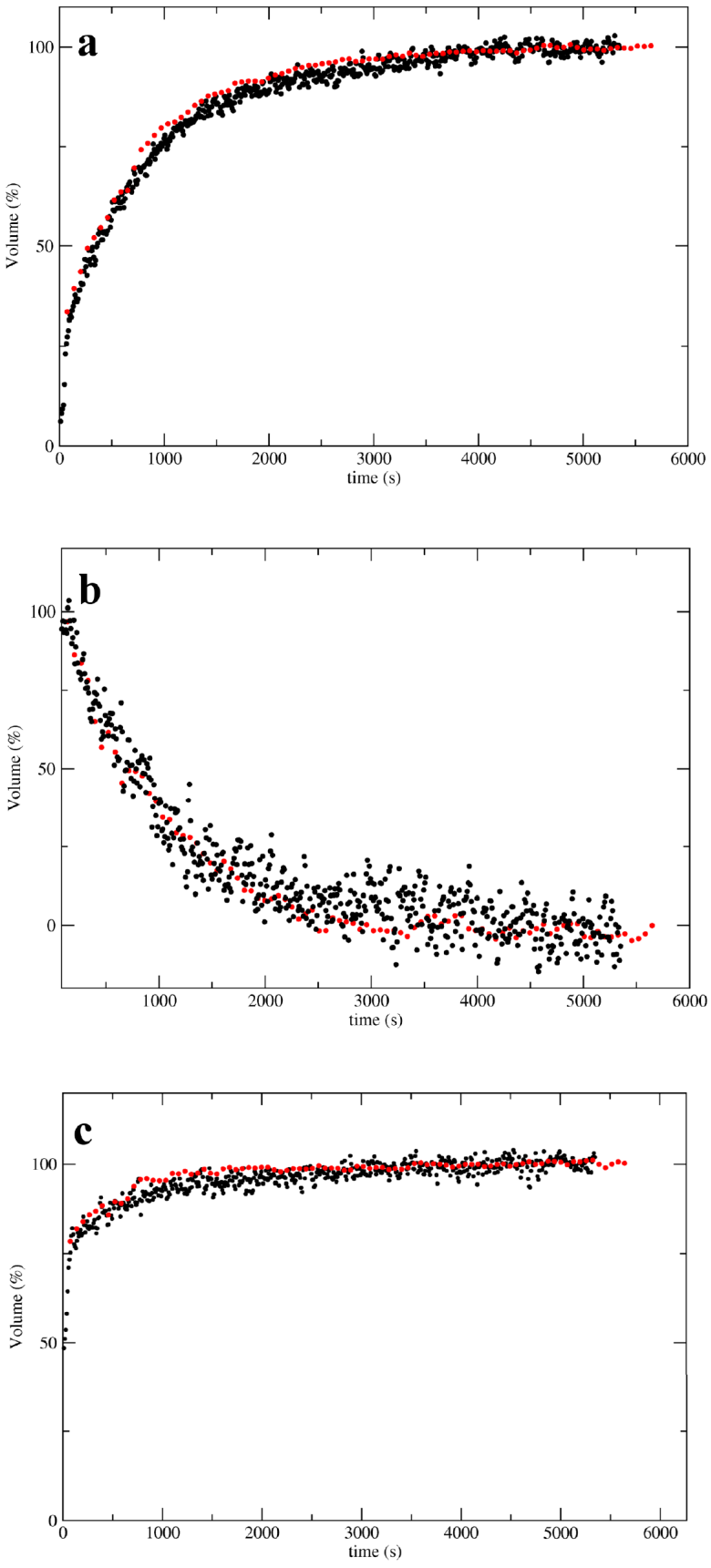


Figure 8.4: Evaluation of the influence of the 1s recycling delay d1 introduced inside the SOFAST pulse sequence (red) on the evolution of the three classes of peaks N (a), I (b) and I+N (c), compared to the signal obtained with the usual shorter d1 of 0.2s (black).

intermediates on parallel pathways have to be discarded. Five states are required to describe adequately the observed kinetics data. Different models can be proposed, all including three long-lived intermediates, all of them NMR invisible but one. Figure 8.5 presents one of the five-state models, but the obtained data are not sufficient to discriminate between the different possibilities.

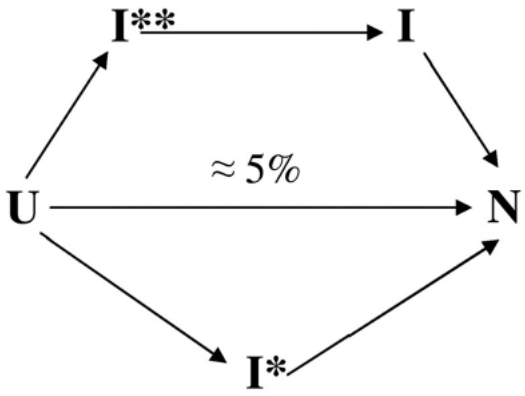


Figure 8.5: Five state folding model for B2M in a phosphate buffer at 25°C, derived from statistical analysis of the folding data obtained in NMR. Stars signal NMR-invisible states.

8.2 Following the folding with methyl protons as probes

As previously seen in section 7.2.3, the SOFAST-HMQC can be adapted to use methyl protons as probes instead of amide protons. This can be of particular interest for two reasons. The first is the location of the methyl protons inside the protein, that is different from that of amide protons. Methyl groups are located on the side chain of amino-acids, while amide protons are located on the backbone of the protein. While methyl groups will usually fold into the hydrophobic core of the protein, amide groups, being in the backbone, can be either exposed to or hidden from the solvent. The second reason that makes the use of methyl protons as probes being the increased sensitivity, as there are three protons for each methyl group. This additional sensitivity can be handy when dealing with lower protein concentrations. Methyl groups having a specific location in the structure of the protein, they may not exhibit the same folding kinetic time constants as amide protons, as the hydrophobic core may require further packing after the backbone has reached its global conformation.

To study this possibility, the folding of B2M at a concentration of 200µM was followed acquiring alternatively an amide SOFAST spectrum and a methyl SOFAST spectrum. Results are shown on figure 8.6. No difference can be observed between the folding rates of amide and methyl protons, but the gain in sensitivity is clear. Yet surprisingly

the signal for the N class for the methyl protons seem to have a small bi-exponential component.

8.3 Folding of the W60G-B2M mutant

W60G is a single point mutant known to have a reduced conformational flexibility, and a lower propensity to form amyloid fibrils, than that of the wild type B2M [59]. Folding studies of this mutant using UV-Fluo spectroscopy revealed a strong difference to the folding of the wild type protein, as W60G folds within seconds, and the slow folding phase that is characteristic of the folding of WT B2M is absent [59]. Yet when W60G refolding is followed using SOFAST-HMQC, the build up of the N state signal shows a clear mono-exponential slow phase, and a burst phase of ∼8%. This difference in the UV-Fluo and the NMR results points towards similar environments of W95 in the intermediate and native states, because fluorescence of B2M mainly comes from W60 and W95, making intermediate and native states of W60G indistinguishable using UV-Fluo spectroscopy. The fact that the I+N signal is not constant either for W60G (figure 8.8) implies the existence of a NMR invisible intermediate state. Yet because the kinetic build-ups for I+N and N signals can be fitted with mono-exponential functions, a simpler four-state model can be proposed for the folding of W60G (figure 8.7). The population of the NMR invisible states is lower in W60G folding (15%, vs 34% in WT folding), which correlates with the higher thermodynamic stability of the mutant protein and its lower propensity to form amyloid fibrils.

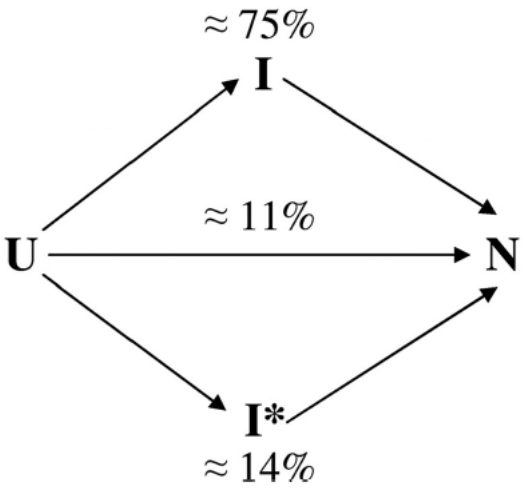


Figure 8.7: Four state model for W60G-B2M folding in a phosphate buffer at 25°C. I* is NMR-invisible. Percentages correspond to the proportion of unfolded proteins using the corresponding pathways to fold.

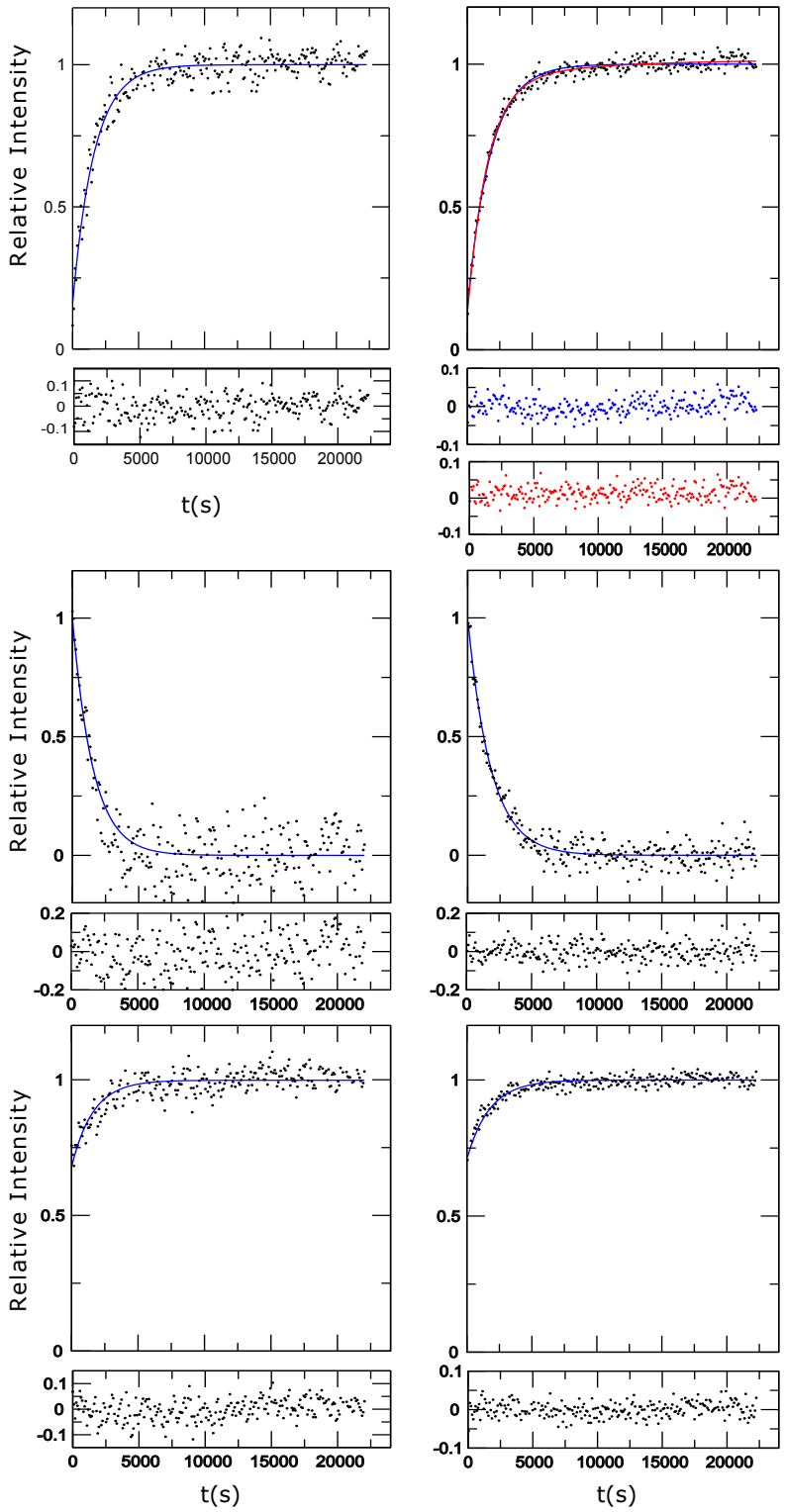


Figure 8.6: Folding curves corresponding to standard (amide) SOFAST (left) and C-H (methyl) SOFAST (right). On both sides, top line corresponds to the native state class, middle line to the intermediate class, bottom line to the I+N class. In each class, the intensities of the peaks are summed, and then normalized to the final or initial value. Data is fitted using mono- or bi-exponential functions depending on the evaluation of the residuals, that are shown below each graph.

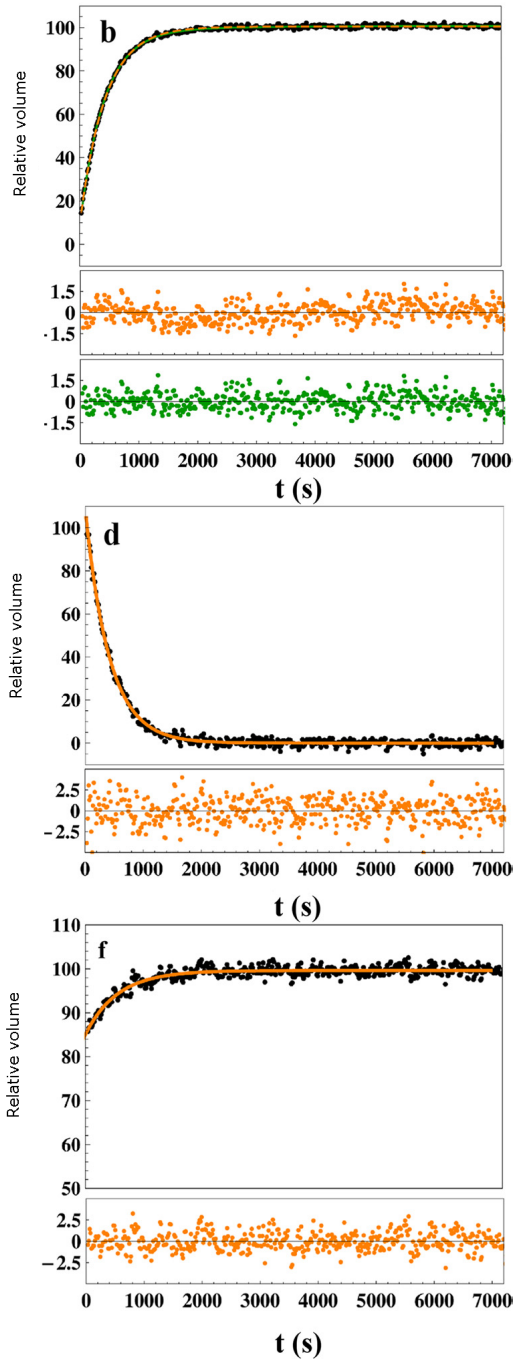


Figure 8.8: Folding of W60G in a Phosphate buffer at 25°C using peak class N (b), I (d), I+N (f). The volume of the peaks is summed inside each class and is normalized to the final or initial value (black); residuals corresponding to mono-exponential (orange) and bi-exponential (green) fitting functions are shown. Data is obtained using the SOFAST pulse scheme, as previously described.

8.4 Conclusions

To conclude on this part, both WT B2M and W60G-B2M populate NMR-invisible, thus structurally native-unlike, folding intermediate states with levels that are far from negligible. The population levels of these NMR-invisible intermediates correlate with the propensity of the protein to form amyloid fibrils. The folding mechanism of WT B2M is likely to include five states (unfolded, one NMR visible intermediate, two NMR invisible intermediates, native). The structure of the NMR visible intermediate is clearly native-like, and the isomerization of the His31-Pro32 peptide bond is the major event of B2M folding.

- The problem of the NMR-invisibility of two of the intermediate states

While the isomerization of P32 had been pointed out through the use of traditional biophysical methods (such as UV-fluo and Circular dichroism) as the main phenomenon occurring during the folding of B2M, NMR studies show that a whole region of the protein is not affected by this isomerization. Along the detection of intermediate states that were not distinguishable by other methods, this is a main illustration of the potential of NMR to bring structural insights to the study of short-lived species, especially intermediate states.

This potential is of course lowered by the inherent limits of the technique: two of the three intermediate states derived from the NMR data are not directly observable in NMR. Two main arguments can explain the NMR-invisibility of these states: either they are oligomeric states, too large to be detected in NMR, or the dynamics of these states imply a peak line-width that is also too large to be detected. Should NMR be discarded as a method to study these invisible states? The fact that they could be indirectly detected suggests that information on these states may be obtained indirectly using NMR.

One strategy is based on one of the explanations for the NMR-invisibility of the intermediates states. If NMR-invisible intermediate states are oligomeric, changes in the experimental conditions should imply detectable changes in the detected NMR signals. Indeed, if conditions that favor the formation of oligomers are used, the population of these invisible states, that is measurable in NMR, should be impacted.

- Comparison of the folding models obtained using NMR or UV-fluo techniques

In order to compare our NMR results with the folding studies based on UV-fluorescence spectroscopy, the probes at the origin of the signal in the two methods must first be considered. NMR methods used here use amide protons as probes, corresponding to 97 probes along the amino-acid backbone of the protein, while UV-fluorescence focuses on the side chain of the two tryptophan residues W60 and W95 (displayed

on figure 4.4(b)). As shown by studying the folding of mutant W60G in UV-fluo, only a fast phase can be detected during the folding, which implies that W95 is affected by the folding only during the transition from the unfolded state to the intermediate states. All intermediate states, and the native states, have a similar micro-environment concerning W95 fluorescence. Considering the folding of the wild-type protein in UV-fluo, and as such the contribution of W60, all intermediate states have a similar micro-environment, but are different both from the native and unfolded state.

While the folding of W60G could be considered as a simple two-state transition directly from the unfolded to the native state based on the UV-fluo data, NMR experiments show that at least two intermediates are necessary to account for the kinetics observed. One of them is directly visible in NMR, the other being NMR-invisible, while the two are invisible in UV-fluo. The folding of the wild-type protein shows a similar difference between the two techniques: while one [143] (or two considering the earlier studies[37, 92]) intermediate states are detected in UV-fluo, three intermediate states are necessary to model the evolution of NMR signals.

These differences in the detection of intermediate states, due to the difference in the probes that are used in the two techniques, makes the evaluation and comparison of different folding models a difficult task. A first order analysis could be that all intermediate states detected in NMR are just populating an UV-fluo indistinguishable ensemble of states that would replace in the model the intermediate states detected in UV-fluo. Yet it is more difficult to account for the bi-exponential slow phase folding detected in NMR, while all kinetics detected in UV-fluo for the slow phase are mono-exponential. This bi-exponential behavior implies the parallel folding pathways of the NMR-based folding models.

A parallel folding pathway implies that the population of the native state is increasing from two different sources (intermediate states), with two different kinetics. One hypothesis can be made to explain the non-detection of these two kinetics in UV-fluo: that one of the intermediate states is indistinguishable from the native state considering the UV-fluo signal, which means that their W60 micro-environment is similar. This would imply that the main event on this pathway would not be the isomerization of P32, but a rearrangement that would not affect either W95 or W60. This rearrangement would have been detectable in NMR, had not the intermediate state concerned been NMR-invisible. The different probes used by NMR and UV-fluo explain the fact that models that seem clearly different are not in principle incompatible. Obtaining more information on these NMR-invisible states may lead to more precise folding models.

Chapter 9

Exploring B2M folding

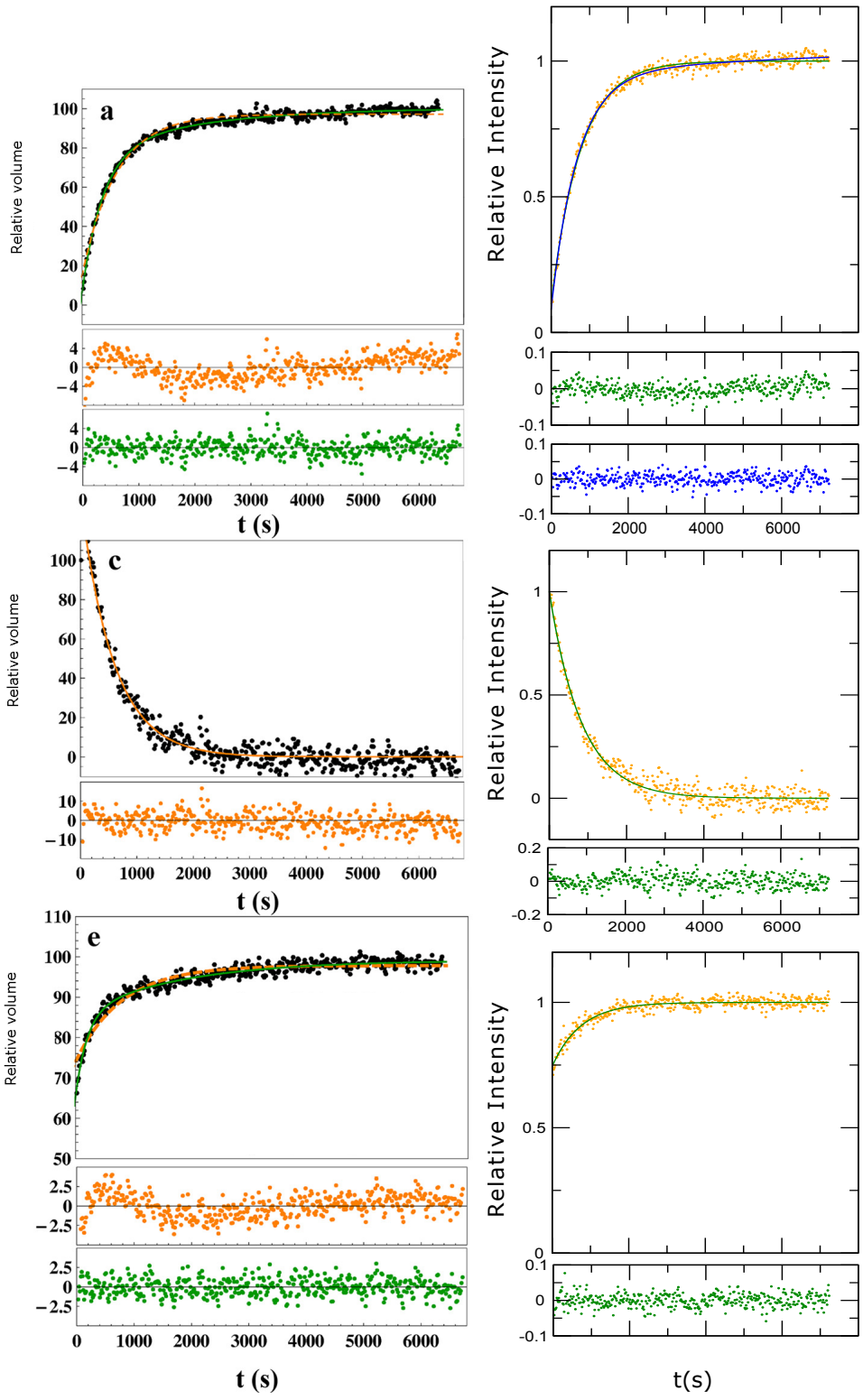
It was shown in chapter 7 that following the folding of a protein in real time using NMR gives access to the folding mechanism, even if the measurements are made only in one set of conditions (buffer, temperature, ...). Yet the protein folding free energy landscape is thought to be a rugged landscape, sensitive to the evolution of thermodynamic or environmental factors. Different states may be populated at different temperatures, or their populations may vary. Real-time NMR can be used to study the robustness of a folding model obtained in one set of conditions, exploring the evolutions of the folding. In this chapter, the influence of two factors on the folding of B2M is investigated: buffer conditions and temperature.

9.1 B2M folding: solvent dependence

Two refolding buffers were used in these B2M folding NMR studies. One was Phosphate 300mM pH=7.2, the other was HEPES 900mM pH=7.2. The reason for the use of a phosphate buffer is that it is the most common buffer in B2M folding studies using other spectroscopic methods, such as UV-Fluo (see 6.2). To be able to compare NMR experiments to UV-fluo experiments among others, a phosphate buffer was used in the first experiments. But phosphate buffers have a drawback in terms of NMR sensitivity, due to their high conductivity [100], when using NMR spectrometers equipped with cryogenic probes. The HEPES buffer, even with its three fold higher concentration, has a conductivity that is 100 fold lower than that of phosphate buffer. As sensitivity is an issue when concentration decreases, the HEPES buffer was used in all following experiments.

To assess for an influence of the buffer on B2M folding, the NMR folding experiments at 25°C were repeated with a HEPES buffer. The comparison of the two folding experiments is shown in figure 9.1. The first result is that for the I+N class, the

signal is fitted with a bi-exponential function for the phosphate buffer, while it is fitted with a mono-exponential function for the HEPES buffer. More dramatic, the folding is globally slower in HEPES ($1.3.10^{-3}s^{-1}$) than in phosphate ($1.7.10^{-3}s^{-1}$) at 25°C. Those differences may be explained by the higher viscosity of the HEPES buffer, due to its higher concentration, the choice of which has anecdotal causes, but further work would be needed to really assess for the causes of these differences.



(a) Folding in Phosphate buffer

(b) Folding in HEPES buffer

Figure 9.1: Comparison of the folding of B2M in two different buffers, using peak class N (top), I (middle), I+N (bottom). As previously described, the volume of the peaks is summed inside each class and is normalized to the final or initial value (black for Phosphate buffer, orange for HEPES buffer); residuals corresponding to mono-exponential (top line) and bi-exponential (bottom line) fitting functions are shown under each graph.

9.2 B2M folding: temperature dependence

The same experimental conditions, ie with a HEPES refolding buffer, can be used to study B2M folding dependence on temperature. Figure 9.3 shows the evolution of the signal for the different classes of peaks (N, I, I+N) at three different temperatures: 15°C, 20°C and the already presented 25°C.

The evolution for the N state class of peaks as a function of temperature shows no impact of temperature on the global aspect of the curves: all have to be fitted with bi-exponential functions. The kinetic time constants (table 9.1) seem to double with every five degrees increase. The burst-phase (table 9.2) appears to be lower at 20°C, but no coherent evolution along the temperature increase arises, staying around the 10% observed at 25°C. The same kind of temperature dependence appears for the I state class, as all used fit functions are mono-exponential, and the same doubling of the time constant for each 5°C increase is observed.

	15°C	20°C	25°C
N	$(3.4 \pm 0.07) \cdot 10^{-4} s^{-1}$	$(6.1 \pm 0.2) \cdot 10^{-4} s^{-1}$	$(1.3 \pm 0.03) \cdot 10^{-3} s^{-1}$
I	$(3.2 \pm 0.1) \cdot 10^{-4} s^{-1}$	$(6.4 \pm 0.3) \cdot 10^{-4} s^{-1}$	$(1.2 \pm 0.05) \cdot 10^{-3} s^{-1}$
I+N	$(3.3 \pm 0.1) \cdot 10^{-4} s^{-1}$	$(6.2 \pm 0.5) \cdot 10^{-4} s^{-1}$	$(1.3 \pm 0.1) \cdot 10^{-3} s^{-1}$

Table 9.1: Temperature dependence of the kinetic time constants obtained using mono-exponential functions of time

	15°C	20°C	25°C
N	$(87.5 \pm 0.9)\%$	$(84.6 \pm 1.0)\%$	$(89.8 \pm 1.1)\%$
I+N	$(39.3 \pm 0.7)\%$	$(32.1 \pm 1.2)\%$	$(25.5 \pm 1.1)\%$

Table 9.2: Burst phases at a concentration of 750µM

The evolution of the I+N class along temperature changes shows more differences. First, as was seen in the solvent dependence, folding curves at 25°C and 20°C can be fitted with mono-exponential functions. Yet surprisingly the I+N curve at 15°C reveals a bi-exponential function of time. As was shown before, a bi-exponential evolution implies a more complex folding mechanism. There seems to be one more intermediate state at 15°C than at 25°C or 20°C, the mechanism at this temperature could then resemble that observed in phosphate buffer.

Furthermore, while the same evolution of the time constants as in the I and N classes is observed, the burst phase has a strong temperature dependence, going from 25% at 25°C up to 39% at 15°C. The amplitude of the burst phase being directly related to the population of the non-native intermediate state, this population increases as temperature goes down. Two hypotheses, based on the existence of NMR invisible intermediate states, were proposed to explain the lack of signal at the beginning of the folding: either those intermediate states are unfolded states, or ensembles of unfolded

states, or they are constituted of oligomeric states or ensembles of oligomeric states. Since the population of unfolded states should rise with the increase in temperature, all other parameters equal, this hypothesis should be discarded. On the contrary, an increase in the population of the NMR-invisible states with decreasing temperature is consistent with those states being oligomeric. The temperature dependence of B2M folding points towards the existence of oligomeric intermediate state(s).

Finally, based on all the kinetic time constants measured, that can be reduced to a single time constant for any given temperature, the energy of activation of the global folding process can be evaluated using equation 9.1 and the temperature dependence of the folding rate. Of course with only three different temperatures, the result must only be considered as an order of magnitude. The obtained value, of the order of magnitude of 90kJ/mol, is also consistent with the cis-trans isomerization of a proline [34].

$$k = A.\exp\frac{-E_a}{R.T} \tag{9.1}$$

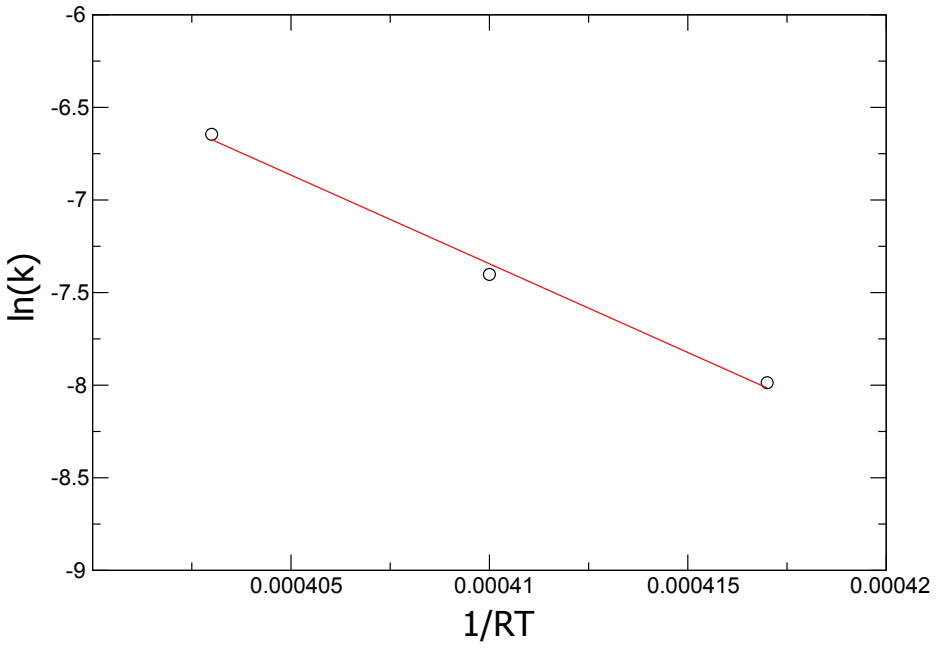


Figure 9.2: Extraction of the activation energy using Arrhenius equation, $ln(k) = f(\frac{1}{R.T})$ at a concentration of 750mM. Linear regression (red line) gives Ea ~95kJ/mol

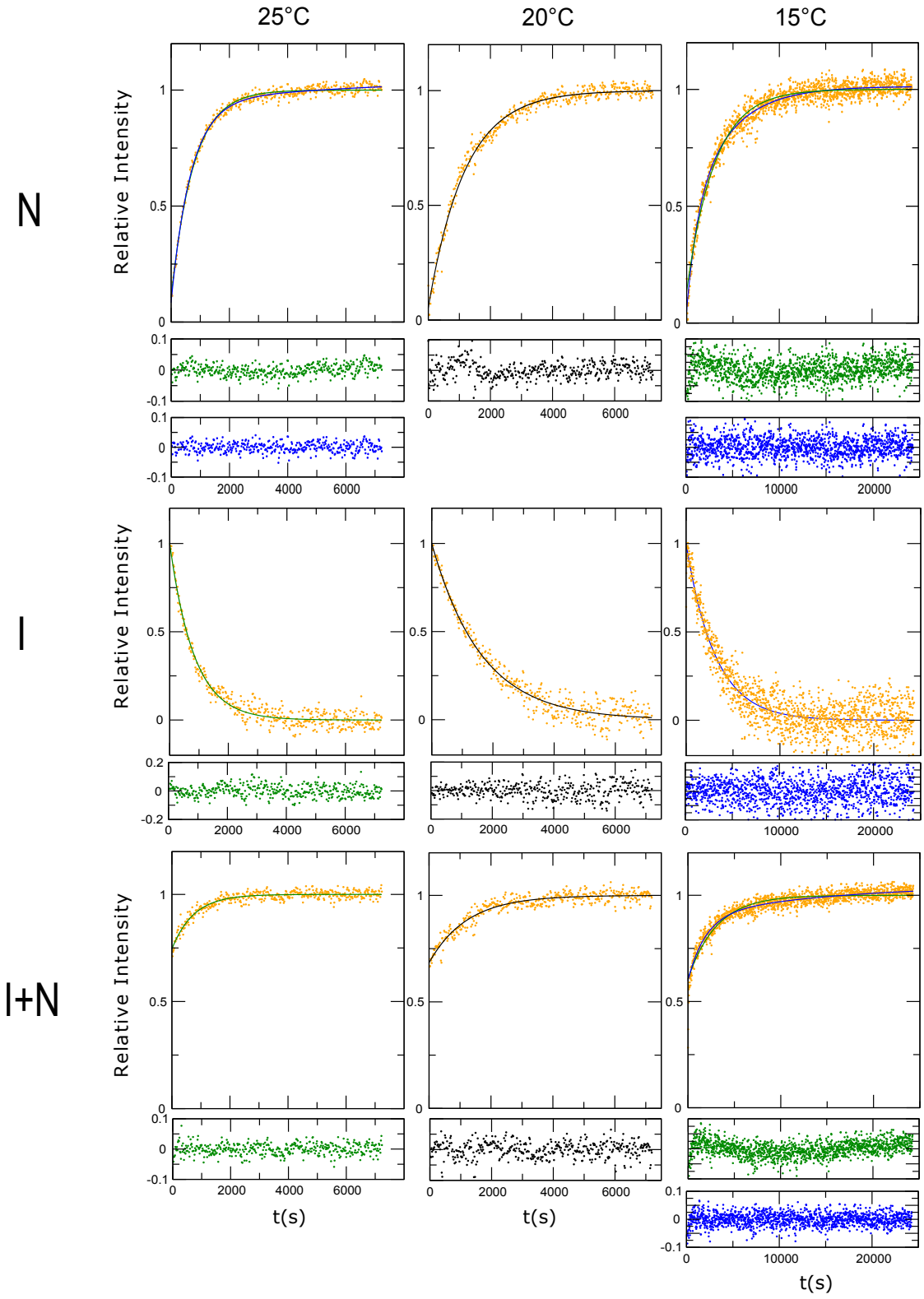


Figure 9.3: Folding of B2M measured at three temperatures: 25°C (left), 20°C (center), 15°C (right), using peak class N (top), I (middle), I+N (bottom). As previously described, the intensity of the peaks is summed inside each class and is normalized to the final or initial value (orange); residuals corresponding to mono-exponential (green or black) and bi-exponential (blue) fitting functions are shown under each graph.

9.3 Fitting the folding signals: mono or bi-exponential functions?

NMR signals for classes N and I+N have different aspects depending on temperature. For a protein concentration of 750µM, N class kinetic traces are fitted with a bi-exponential function at 25°C and 15°C, but a mono-exponential function at 20°C. Since the noise level is roughly the same, it cannot explain the difference of aspect of the curves. Another example is the aspect of the I+N curve, at a protein concentration of 750µM, that is mono-exponential at 25°C and 20°C, but bi-exponential at 15°C.

These differences of aspect reveal a very complex folding mechanism. Are mechanisms totally different depending on the temperature, for example simpler at 20°C, as in this case all curves can be fitted with mono-exponentials? Or at 20°C are we witnessing a crossing in the evolution of the kinetic rates, due to different activation energies? For example, the phenomenon corresponding to the faster kinetic rate at 25°C could correspond to the slower at 15°C, both rates being equal at 20°C while the folding mechanism remains just as complex at all temperatures. The complexity being too high, one must consider reducing the model to only mono-exponential kinetic curve fittings, and those are the fittings used to extract kinetic time constants and burst phases as displayed in tables 9.1 and 9.2.

9.4 B2M folding: Simplified model

Building a model for the folding of a protein can be very complex. With only one condition, 750µM, 25°C, phosphate buffer, five states are already needed to comply to the observed kinetic parameters. When globally looking the conditions in which the folding of B2M was followed, the picture can seem senseless. For example the folding follows bi-exponential kinetics for the apparition of the N state at 25°C and 15°C, but not at the intermediate temperature of 20°C. Solvent and temperature have effects on the folding mechanism of B2M. In order to extract some sense from all this data, the model must be simplified. This is why only a mono-exponential fit was used to extract the kinetic parameters that will be used to build a folding model. Two characteristics of the folding are essential in the building of the model:

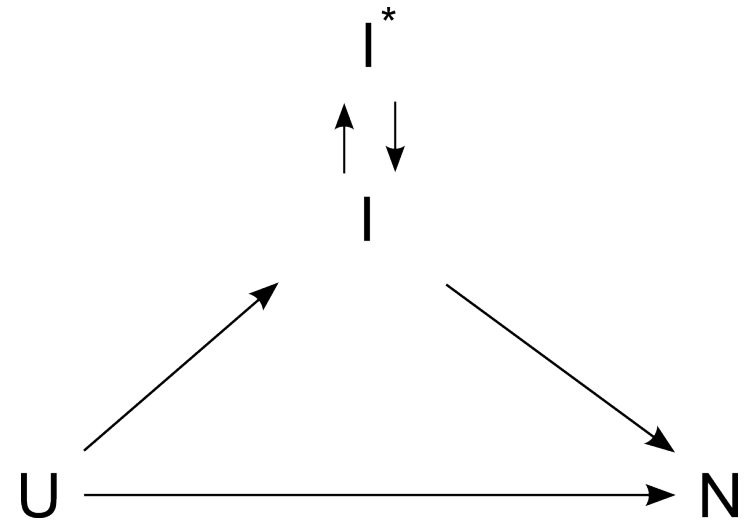
- the folding intermediates can be divided into two categories, one is the NMR visible intermediate state that was identified before (termed I), the other being the NMR invisible state, that may be an ensemble of states, and is expected to be oligomeric (termed I*).
- the folding characteristic time constants can be reduced to a unique time constant corresponding to the isomerization of the H31-P32 peptide bond from the

trans to the cis conformation, implying that all folding intermediate states are in a trans conformation.

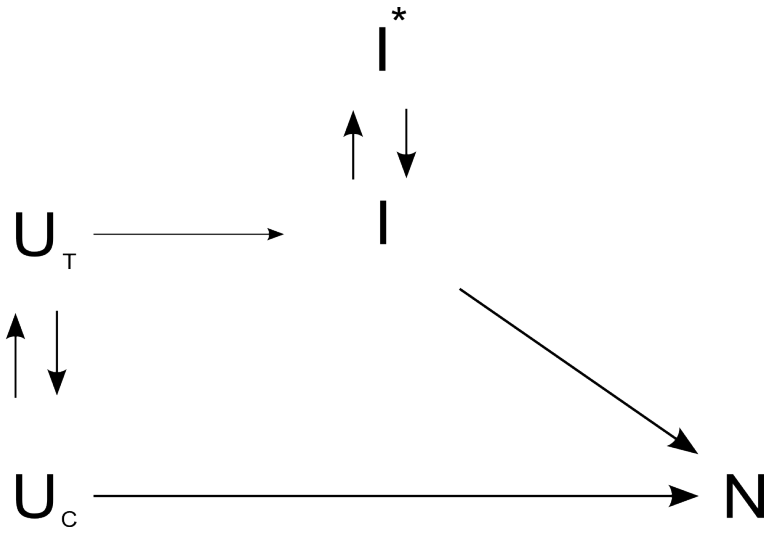
Finally, the burst phase on the build-up of the N class signal means necessarily that circa 10% of the population of the protein folds within the dead-time of these NMR experiments directly to the native state. From these considerations, a simplified model can be proposed, relevant to the real-time NMR timescale. The model presented on figure 9.4 consists of one unfolded state (U), that folds either directly to the native state (N) or to the NMR-visible intermediate state I, both with kinetics that are considered fast relative to the isomerization rate of the proline, and to the dead time of the NMR experiments. An equilibrium between I and I* appears at a rate that again is fast relative to both the isomerization rate and the dead time of the experiments, considering that I* is fully populated in the first spectrum of the NMR experiments. The fact that proteins in a trans conformation for the proline can only fold to the native state through the I state explains the unique slow kinetic time constant that is observed. More complex models could be built based on the same data, and this is clearly a simplification giving up the bi-exponential characteristics of certain folding kinetic curves. This simplified model can be complemented with details on the unfolded state, that may be split into two states in slow equilibrium corresponding to the two possible conformations for the H31-Pro32 peptide bond [37, 92, 143], taking into account the fast folding phase observed in UV-fluo.

These simplified models are clearly compatible with models presented in section 6.2, the equilibrium I - I* replacing in the models the unique I_t intermediate state, which is not problematic as long as I and I* are indistinguishable in UV-fluo.

Studying the evolution of protein folding can be achieved using real time NMR and appropriate thermodynamic or environmental changes. In the case of B2M, folding is clearly dependent on those factors, which can lead to unresolvable complexity. Yet through careful modeling, insights such as the isomerization of the proline being the main folding event, or the fact that the two intermediate states are probably in equilibrium, can be obtained.



(a)



(b)

Figure 9.4: Folding models derived from the NMR folding data obtained at different temperatures, with and without the distinction of two unfolded states. I* is NMR-invisible.

Part IV

Probing the NMR invisible state

Part III established that B2M populates a NMR-invisible state during its folding, and that this invisible state is probably oligomeric. This part aims at the characterization of this NMR-invisible state, first through the indirect means of NMR spectroscopy, then with other biophysical methods such as Small Angle X-ray Scattering or infrared spectroscopy, and finally test the P32A mutant of B2M as an in-equilibrium model for this transient NMR-invisible state.

Chapter 10

Detecting a transient, invisible oligomer using real time NMR

Probing NMR-invisible states using NMR can seem counter-intuitive. Yet just as a lack of signal could be interpreted as an indirect proof of existence of this invisible intermediate state, NMR can be used to characterize it. Especially, following the folding in NMR is a mean to test the hypothesis that the invisible state is an oligomeric state, using the influence of concentration.

10.1 The influence of concentration on the folding of B2M

Five different final concentrations were used to study B2M folding dependence on protein concentration, using two different atomic probes: 750µM, 375µM, 188µM using amide protons as probes, and 150µM and 50µM using methyl protons as probes. Figures 10.1 and 10.2 displays the results of the folding at these concentrations at 20°C. Samples were prepared as previously stated, and diluted in the Urea buffer to reach the desired concentration.

All N class signals show mono-exponential evolutions in these conditions. Concentration doesn't seem to influence the mechanism of folding of B2M in that matter. For all concentrations, at 20°C, the kinetic time constants seem to be similar ($\sim 6.10^{-4}$ s^{-1}), and the burst phase is around 10%. Concentration doesn't seem to be of influence for the apparition of the native state. In a similar fashion, concentration doesn't influence the decay of the I class signals, and time constants are similar to those of the N class.

But interestingly, when considering the I+N class, the population of the NMR-invisible state(s) decreases with the decrease of protein concentration, from 32% at 750µM to 19% at 188µM, and the I+N signal is even constant at the lowest concentration considered, 50µM. These results are clearly consistent with NMR-invisible states being an oligomeric state or ensemble of states.

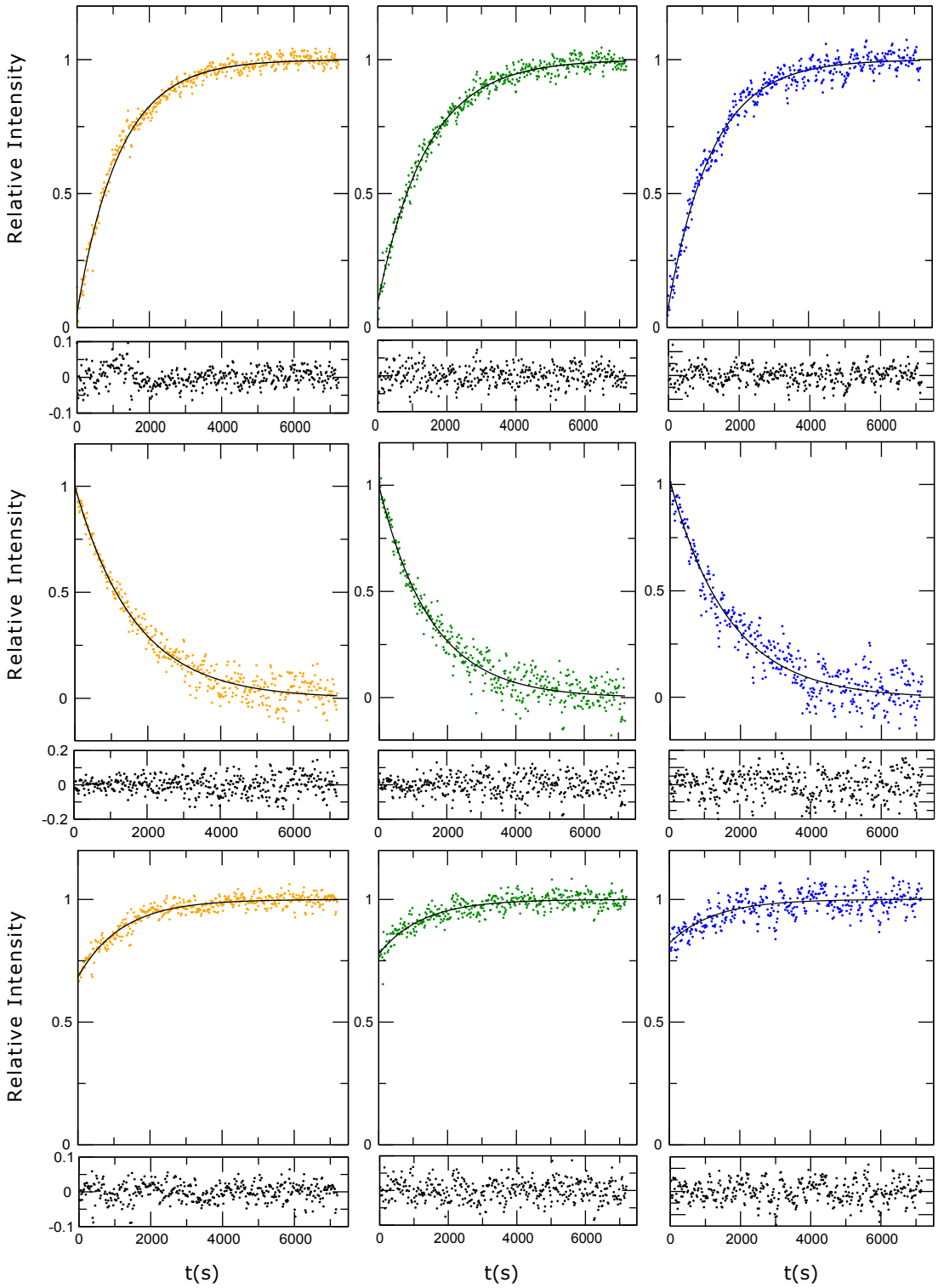


Figure 10.1: Folding of B2M followed at three different concentrations: 750µM (yellow, left), 375µM (green, center), 190µM (blue, right), using peak class N (top), I (middle), I+N (bottom). As previously described, the intensity of the peaks is summed inside each class and is normalized to the final or initial value; residuals corresponding to mono-exponential (black) fitting functions are shown under each graph.

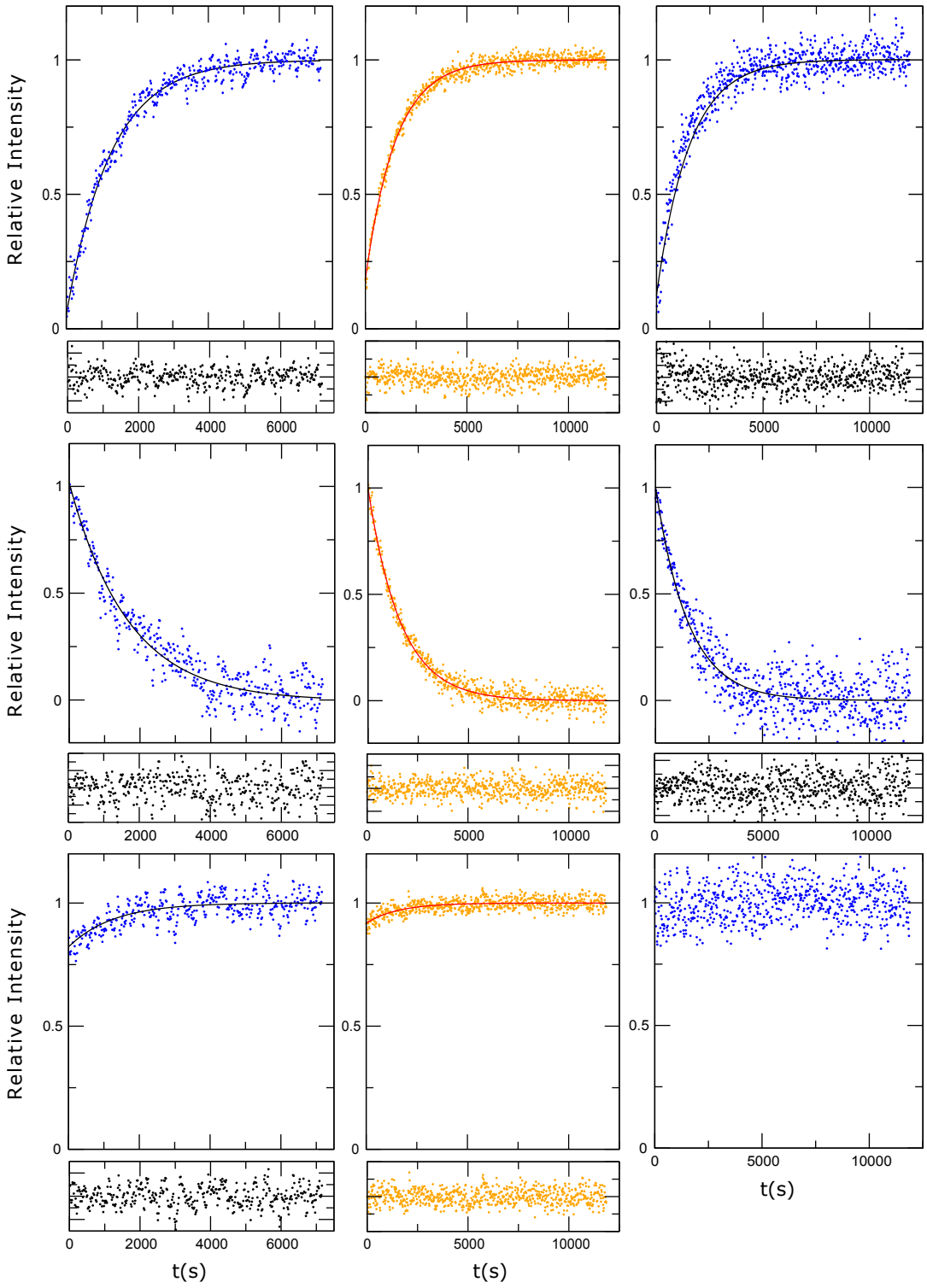


Figure 10.2: Folding of B2M followed at three different concentrations: 190µM (left), 150µM (center), 50µM (right) , using peak class N (top), I (middle), I+N (bottom). As previously described, the intensity of the peaks is summed inside each class and is normalized to the final or initial value; residuals corresponding to mono-exponential fitting functions are shown under each graph.

10.2 B2M folding: discussion on both temperature and concentration influence

To get a better knowledge of the B2M folding energy landscape, the dependence on temperature and concentration, and to increase the robustness of the conclusions drawn, three concentrations (750µM, 375µM, 188µM) were studied at the three temperatures (25°C, 20°C, 15°C), giving 11 conditions with the two lower concentrations studied using Met-SOFAST-HMQC at 20°C. Results are presented in table 10.1 for the kinetic time constants, in table 10.2 for the burst phase parameters for classes N and I+N, and on figures 14.1, 14.2, 14.3 and 14.4 in annex VI. A first conclusion that can be drawn is that N, I and I+N classes have similar kinetic time constants in any given temperature-concentration conditions. This supposes that the folding is mediated by only one time constant. Of course this is a simplification, as certain curves, as previously stated, should be fitted with bi-exponential functions of time. Yet this can be considered as a first order analysis. The unique kinetic time constant corresponds to that of a proline cis-trans isomerization. This folding rate does not depend on concentration at any given temperature.

Class	T	750µM	375µM	188µM
N	15°C	$(3.4 \pm 0.07) \cdot 10^{-4} s^{-1}$	$(4.1 \pm 0.1) \cdot 10^{-4} s^{-1}$	$(4.1 \pm 0.2) \cdot 10^{-4} s^{-1}$
	20°C	$(6.1 \pm 0.2) \cdot 10^{-4} s^{-1}$	$(6.4 \pm 0.2) \cdot 10^{-4} s^{-1}$	$(6.2 \pm 0.3) \cdot 10^{-4} s^{-1}$
	25°C	$(1.3 \pm 0.03) \cdot 10^{-3} s^{-1}$	$(1.2 \pm 0.07) \cdot 10^{-3} s^{-1}$	$(1.3 \pm 0.06) \cdot 10^{-3} s^{-1}$
I	15°C	$(3.2 \pm 0.1) \cdot 10^{-4} s^{-1}$	$(3.8 \pm 0.2) \cdot 10^{-4} s^{-1}$	$(3.6 \pm 0.4) \cdot 10^{-4} s^{-1}$
	20°C	$(6.4 \pm 0.3) \cdot 10^{-4} s^{-1}$	$(5.2 \pm 0.6) \cdot 10^{-4} s^{-1}$	$(4.9 \pm 0.8) \cdot 10^{-4} s^{-1}$
	25°C	$(1.2 \pm 0.05) \cdot 10^{-3} s^{-1}$	$(1.1 \pm 0.08) \cdot 10^{-3} s^{-1}$	$(1.2 \pm 0.1) \cdot 10^{-3} s^{-1}$
I+N	15°C	$(3.3 \pm 0.1) \cdot 10^{-4} s^{-1}$	$(4.2 \pm 0.7) \cdot 10^{-4} s^{-1}$	$(4.6 \pm 1.3) \cdot 10^{-4} s^{-1}$
	20°C	$(6.2 \pm 0.5) \cdot 10^{-4} s^{-1}$	$(7.6 \pm 1.2) \cdot 10^{-4} s^{-1}$	$(6.3 \pm 1.7) \cdot 10^{-4} s^{-1}$
	25°C	$(1.3 \pm 0.1) \cdot 10^{-3} s^{-1}$	$(1.1 \pm 0.2) \cdot 10^{-3} s^{-1}$	$(1.0 \pm 0.8) \cdot 10^{-3} s^{-1}$

Class	T	150µM	50µM
N	15°C	-	-
	20°C	$(6.7 \pm 0.2) \cdot 10^{-4} s^{-1}$	$(6.6 \pm 0.9) \cdot 10^{-4} s^{-1}$
	25°C	-	-
I	15°C	-	-
	20°C	$(6.2 \pm 0.2) \cdot 10^{-4} s^{-1}$	$(6.5 \pm 0.6) \cdot 10^{-4} s^{-1}$
	25°C	-	-
I+N	15°C	-	-
	20°C	$(6.4 \pm 1.1) \cdot 10^{-4} s^{-1}$	not measurable
	25°C	-	-

Table 10.1: Summary of measured kinetic time constants for the folding of B2M

Based on the results displayed in table 10.2, the burst phase for the N class doesn't seem to depend either on concentration or on temperature, with a constant value around 10-15%. This value represents the population of the protein that goes directly from the unfolded state to the native state. This is usually explained by the existence

Class	T	750µM	375µM	188µM	150µM	50µM
N	15°C	$(87.5 \pm 0.9)\%$	$(86.9 \pm 1.4)\%$	$(89.6 \pm 2.1)\%$	-	-
	20°C	$(84.6 \pm 1.0) \cdot \%$	$(82.3 \pm 1.3) \cdot \%$	$(83.7 \pm 2.2) \cdot \%$	$(82.4 \pm 1.0) \cdot \%$	$(88.1 \pm 2.5) \cdot \%$
	25°C	$(89.8 \pm 1.1) \cdot \%$	$(89.2 \pm 3.2) \cdot \%$	$(91.2 \pm 2.7) \cdot \%$	-	-
I+N	15°C	$(39.3 \pm 0.7) \cdot \%$	$(30.2 \pm 1.1)\%$	$(22.8 \pm 2.1)\%$	-	-
	20°C	$(32.1 \pm 1.2) \cdot \%$	$(22.4 \pm 1.4) \cdot \%$	$(12.9 \pm 3.0) \cdot \%$	$(9.2 \pm 2.1) \cdot \%$	~ 0
	25°C	$(25.5 \pm 1.1) \cdot \%$	$(16.7 \pm 2.4) \cdot \%$	$(11.0 \pm 5.0) \cdot \%$	-	-

Table 10.2: Summary of measured exponential parameters (burst phase) for the folding of B2M

of an equilibrium between a cis H31-P32 peptide bond unfolded conformation (U_c) and a trans H31-P32 peptide bond unfolded conformation (U_t). While the protein in a trans conformation must undergo a kinetically unfavourable trans to cis isomerization, the proteins having a peptide bond that is already in a cis conformation in its unfolded state can fold directly to the native state within seconds, as shown by Uv-Fluo experiments. This U_c-U_t equilibrium doesn't seem to be dependent on temperature in the considered range, with a trans conformation being more stable in unfolding conditions while a cis conformation is favoured in the native state.

The major result of this study is of course the dependence of the I+N class burst phase on concentration and temperature. The population of the NMR-invisible state rises both with the increase in concentration and with the decrease in temperature, and goes from undetectable levels at very low concentration up to 40% at 15°C and a concentration of 750µM. This is consistent with the hypothesis of NMR-invisible folding intermediate state(s) being oligomeric. In the following, these oligomeric states will be termed I_2, while the monomeric I state will be termed I_1.

10.3 The oligomeric intermediate state is in equilibrium with the monomeric intermediate

10.3.1 Unfolded state does not show oligomerisation

While the first explanation that comes to mind is that the oligomeric intermediate state is in equilibrium with the monomeric intermediate state, another hypothesis is yet to be considered: this oligomeric state may derive from an oligomer present in the acid-unfolded state, and fold to the native state on a path that would be strictly parallel to that of the monomeric intermediate state.

The oligomerisation of the acid-unfolded state can not be studied using SAXS as in chapters 11 and 12 because of interparticular effects, shown in annex on figure 14.5. Therefore relaxation NMR methods (^1H R_1 and R_2) were used to evaluate the effective local tumbling correlation time τ_c (see equation 7.2) of the acid-unfolded

state at different concentrations along the backbone of the protein. Figure 10.3 shows that the τ_c of the acid-unfolded state is independent of concentration, which means that the hypothesis of an oligomerisation of this state has to be discarded, and that the oligomeric and monomeric intermediate states are therefore in equilibrium.

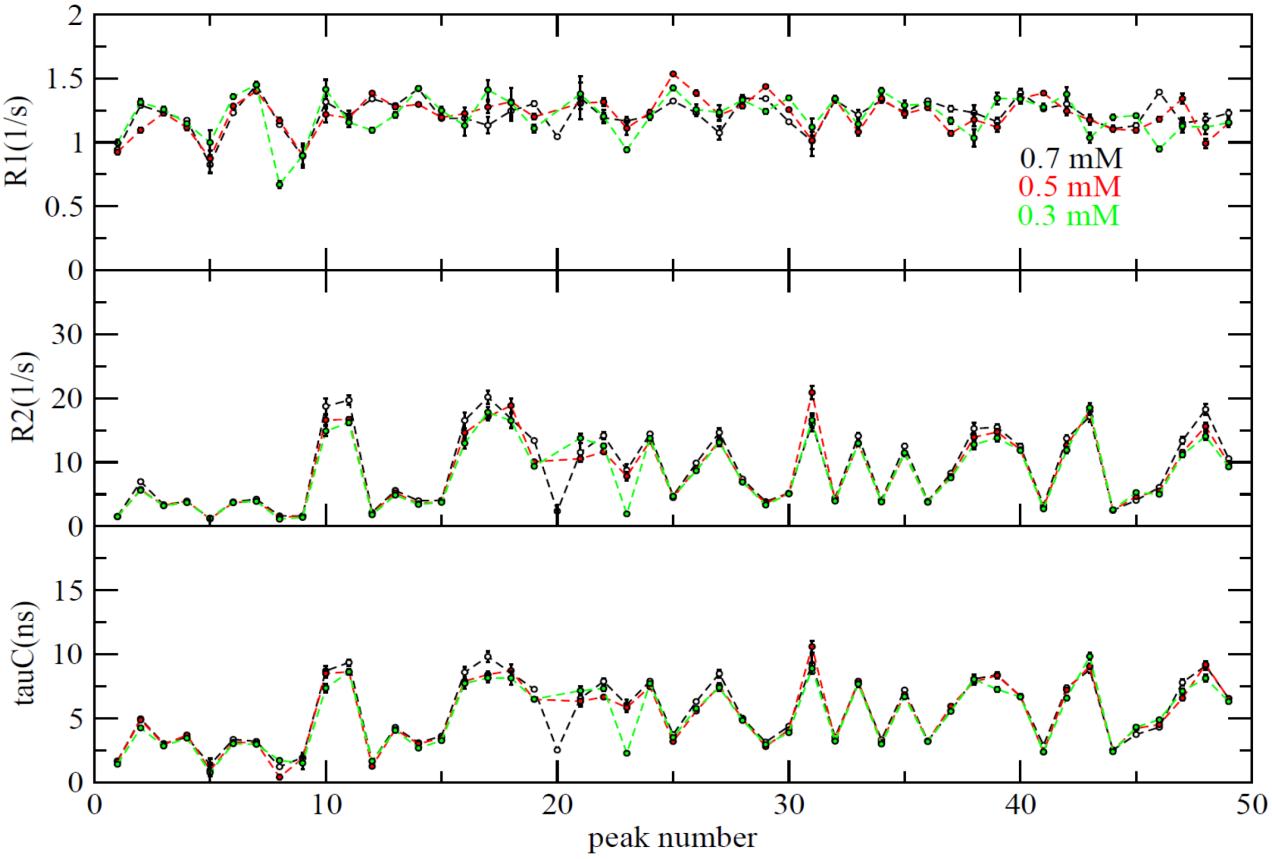


Figure 10.3: Relaxation NMR measurements for the acid-unfolded state of WT B2M, showing that the τ_c is independent of concentration. For methodological description, see 13.4.1

10.3.2 Characterization of the I_1 - I_2 equilibrium

Several hypotheses can be drawn on the I_1 - I_2 equilibrium to account for the data obtained using real time NMR. These data can be resumed in a few points: Only one set of peaks can be detected. The linewidth of the peaks, as measured in the SOFAST-HMQC spectra, is similar to that of the native state. The introduction of a longer recycling delay d1 has no influence on the intensity of the peaks of the detected intermediate state. From these elements, three hypotheses:

- $I_1$and I_2 are in slow exchange. I_1 is monomeric and NMR-visible, while I_2

is oligomeric and invisible. This means that I_2 is a large oligomer, so that the peaks are broaden by its relaxation properties due to a large tumbling correlation time.

- $I_1$and I_2 are in fast exchange. I_1 is monomeric, while I_2 is a low molecular weight oligomer. Therefore the intermediate state spectrum that we see in NMR is a weighted average of the I_1 and I_2 states spectra. But both the linewidth and the T_1 relaxation parameter should be different between this averaged intermediate state and the native state.

- $I_1$and I_2 are in slow exchange. I_1 is monomeric, while I_2 is a low molecular weight oligomer that should be NMR-visible. I_1 and I_2 spectra exactly overlap, so that only one spectrum is visible. But again both the linewidth and the T_1 relaxation parameter should be different between this averaged intermediate state and the native state.

So based on this discussion, the most plausible explanation is that I_2 is a large oligomer in slow exchange with I_1. But considering the low precision of the two criteria retained here (linewidth of a SOFAST-HMQC spectrum and longer d1 recycling delay), the two other hypotheses should not be discarded just yet, and further work should focus on the size of the I_2 oligomeric intermediate state. One last element on this equilibrium, its possible timescale range has an upper limit: since I_2 is fully populated on the first spectrum after the pH jump, the equilibrium has to be faster than the deadtime of the experiment, wich is $\sim 10s$, so the timescale of the equilibrium is probably lower than the second.

Chapter 11

Expanding the scope: Biophysical methods for a further characterization of the oligomeric intermediate

If NMR can only provide indirect evidence of the presence of additional intermediate states, other biophysical techniques may be used for their direct characterization. Since the invisible state that appears during the folding of B2M is an oligomer, Small Angle X-ray Scattering (SAXS) is a technique of choice, as it allows the measurement of the average molecular weight of the sample, as shown in section 3.5. Infrared spectroscopy allows the detection of the secondary structure content of a macromolecule, whatever its size, and may as such provide useful information on the structure of the oligomeric intermediate state. Finally, UV-fluorescence spectroscopy in combination with the use of Thioflavin-T would detect amyloid-like structures inside the oligomeric state, which is relevant considering the energy landscape of B2M presented in section 6.4.

11.1 SAXS: characterizing the stoichiometry of the oligomeric intermediate

Small angle X-ray scattering can provide precise information on the molecular weight, radius of gyration, and if the signal is good enough, the shape, of a protein in solution. Most common conditions for SAXS imply the use of monodisperse samples, because

the signal being additive, the parameters obtained are the average of the parameters of all the molecules inside the sample. Following a kinetic reaction, and as such following the transformation of one molecule into another, implies working on a polydisperse sample. Yet, considering the folding of B2M, the average molecular weight and radius of gyration would be valuable hints, as NMR data suggest the existence of an oligomeric intermediate state. With a population that goes up to 40% at the beginning of the folding, this oligomeric intermediate state(s) should be detectable using SAXS.

11.1.1 Methods

To allow a good comparison, buffer conditions that were used in SAXS are the same as those used in NMR, and experiments were made at room temperature. One SAXS curve is obtained as the average of either 10 illuminations of 10 seconds each, or 20 illuminations of 5 seconds each. The SAXS curve of WT B2M is shown in figure 11.1. It is a curve that is typical of a globular protein, except at the very low angles, where the signs of an aggregation appear. Even though B2M being an amyloid forming protein, and as such aggregation prone, it can be surprising to find aggregates despite the ultra-centrifugation that was used to prepare the sample, following the same protocol as for NMR experiments. As the SAXS signal is the addition of the signals of all the constituents of the sample, and as the larger the object, the fastest the decay of the signal along the growth of the angle of scattering, reliable information can be obtained on the protein that is not aggregated as long as the signal at the very low angles is discarded.

SAXS can also be used to follow the folding of B2M. As the highest population of the oligomeric intermediate state is reached just after the pH jump, the time between the initiation of the folding and the beginning of the measurements, the so called dead time, is important in this experiment. The dead time is linked to the time needed to obtain a homogeneous mixing after the pH jump, and to the time needed to start the measurement after that. Two experimental set-ups have been used to initiate the folding: one using the robotised sample changer available on site (ESRF Grenoble), the other using a by hand mixing. Using the robot has the very interesting advantage to allow the security manoeuvres to be executed before the mixing is induced. Security manoeuvres, which cannot be eliminated, include closing the sample holder, making a security search inside the experimental hutch, and closing the experimental hutch. To be able to avoid this, or at least to make it before starting the folding, can help reducing the dead time. Yet, despite this advantage and the upgrades in the experimental set-up that induced a greatly improved dead time, by-hand mixing was favored, as the robot was not able to successfully and homogeneously mix the Urea-dissolved unfolded protein and the folding buffer, probably due to the viscosity

of the samples. By-hand mixing lead to a dead time of about 50 seconds before the beginning of the first exposure to X-rays.

Once the folding is initiated and the measurements are started, the time resolution is also limited. First because a standard SAXS curve is acquired in 100 seconds, but also, to avoid radiation degradation of the protein, a new sample must be used for each data point along the folding, consisting of 30µL of protein. The amount of protein available being limited, the number of points along the folding, and the time at which the measurements are made, have to be carefully chosen.

11.1.2 Discarding the signal corresponding to the large aggregates

The ability to discard the signal corresponding to the very low scattering angles depends on two conditions. First the population of aggregates must be constant, otherwise the phenomenon observed during the SAXS experiment is not the folding, but the dissociation of the aggregates, or a combination of both. Second condition is that there must be no interaction between the monomers and the aggregates, because for example if the size of the aggregates changes, the relative signals of the aggregates and of the lower molecular weight species will change, making the SAXS data on folding unusable. To examine whether these two conditions are verified, the folding was studied using negative staining electron microscopy. The experiment was made at room temperature. Three measures were made: just after, 15 minutes after, and 3 hours after the initiation of the folding. Buffer conditions are the same as in SAXS and NMR (HEPES buffer). Electron microscopy demands concentration conditions that are different from those used in SAXS and NMR: a maximum of 50 µg/mL, circa 4µM for B2M.

In order to be in conditions that are as close as possible to those previously used, the folding was made at 750µM, and samples were taken for measurement at defined times after the pH jump, diluted 200fold, and dried during the negative staining procedure. The process of dilution and drying takes about 30 seconds, which means that oligomeric states would have evolved to monomeric states. Figure 7.9 display representative pictures of the three samples. It shows that aggregates are present, with a diameter of about 20nm. The population and size of these aggregates seems to be constant, even though electron microscopy is not a precise method for quantitative evaluations, and the evaluated size may be different to that present in solution during folding experiments, due to the 30 seconds evolution time in a diluted environment. In conclusion, discarding the signal corresponding to these aggregates in SAXS seems robust. Another remark is that filtration would not help getting rid of these aggregates, as they are too small.

11.1.3 Assessing SAXS precision using in-equilibrium measurements

Guinier analysis, as explained in section 3.5, can be performed on the SAXS curve of WT B2M and gives an extrapolation of the forward scattering intensity and an estimation of the average radius of gyration of the protein. The forward scattering intensity is directly proportional to the concentration and the molecular weight of the protein in solution. The molecular weight can be calculated by comparison to a standard of known molecular weight (BSA), using the concentrations of the sample and the standard measured with a UV spectrophotometer equipped with a nanodrop device. Results are shown in table 11.1. Measured average molecular weight of the sample is 11.8kDa, radius of gyration (of the protein and its hydration shell) 20.1Å, with robust statistics. Theoretical values are 11.73kDa and 19.06 Angstroms, as calculated using Protparam and Crysol. These equilibrium values are close to the theoretical ones.

	I_0	$R_g(\AA)$	Fidel.	first point used	last point used	$M_w(kDa)$
WT B2M	2,09	20,1	0,98	43	77	11,8
BSA	11,75	30,7	0,91	39	81	66,42

Table 11.1: Guinier analysis of WT B2M in equilibrium

But the calculation of the molecular weight is based on two values, one, the concentration measurement, which takes into account the proteins that are in the ~ 20 nm diameter aggregates, the other, the forward scattering intensity, which doesn't. This leads to an underestimation of the average molecular weight in the sample. As such, the higher than expected molecular weight and radius of gyration may be revealing the existence of a native oligomeric state.

11.1.4 Folding of B2M: two oligomeric states?

As seen in section 11.1.1, the number of measurements during the folding is limited, but carefully chosen, should allow a good evaluation of the phenomena at stake. Figure 11.2 shows the overlay of SAXS curves obtained at different moments during the folding.

As explained above, the first points, at low angles, shown on figure 11.2, have to be discarded to take into account the aggregation. Visually, the forward scattering is decreasing along the folding, as shown by the evolution of the intensity of the signal at lower angles. There is no more evolution after 60 minutes of folding. As the forward scattering intensity is directly related to the average molecular weight, this proves that average molecular weight decreases during the folding of B2M, confirming the existence of an oligomeric folding intermediate state.

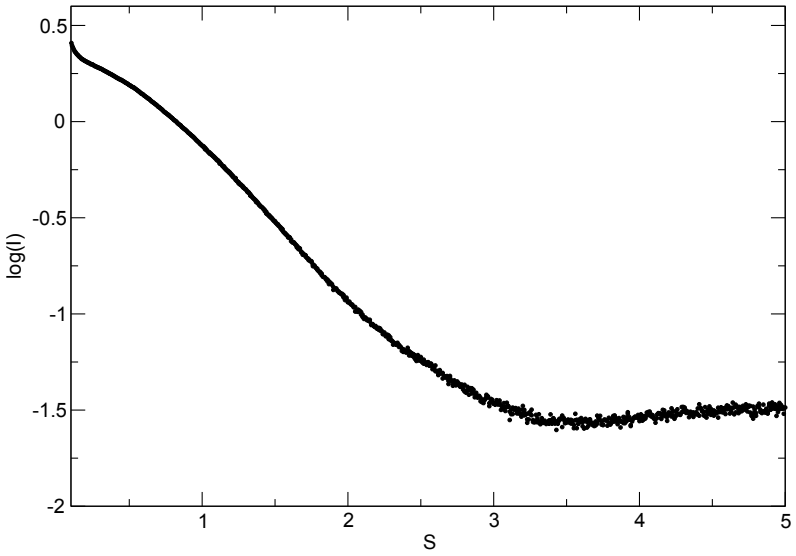


Figure 11.1: SAXS curve of WT B2M. SAXS curve is obtained as the average of 10 illuminations of 10 seconds each, and subtraction of the solvent diffraction curve, using the Bio-SAXS beam-line at the ESRF Synchrotron in Grenoble.

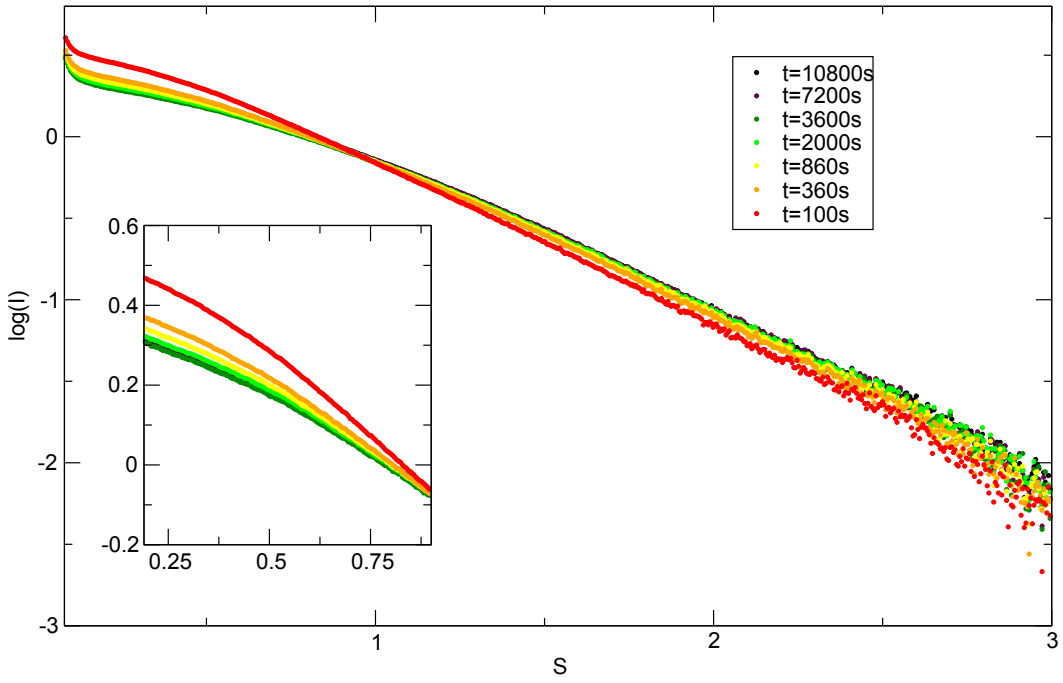
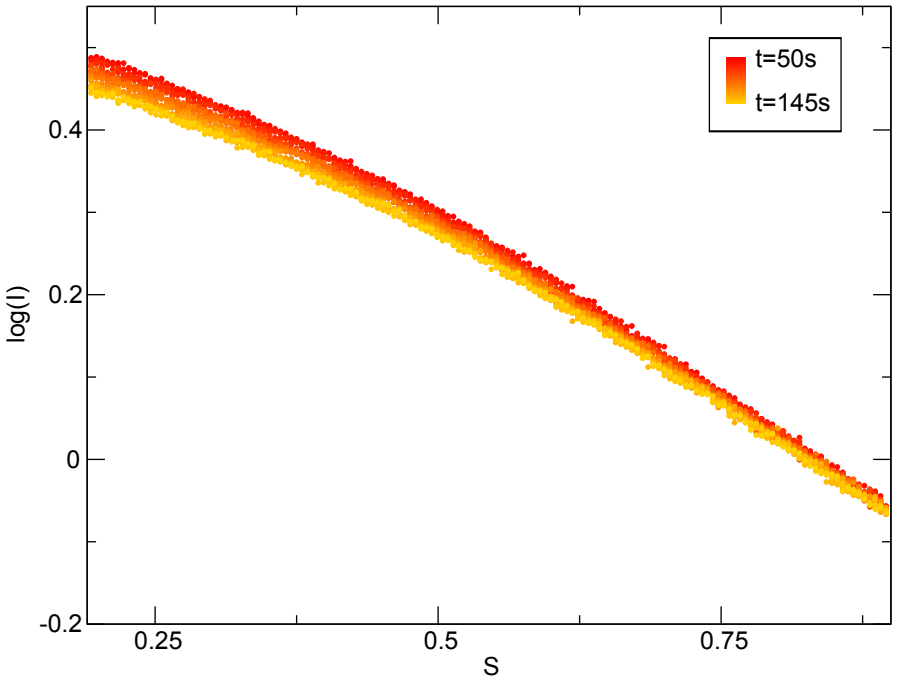
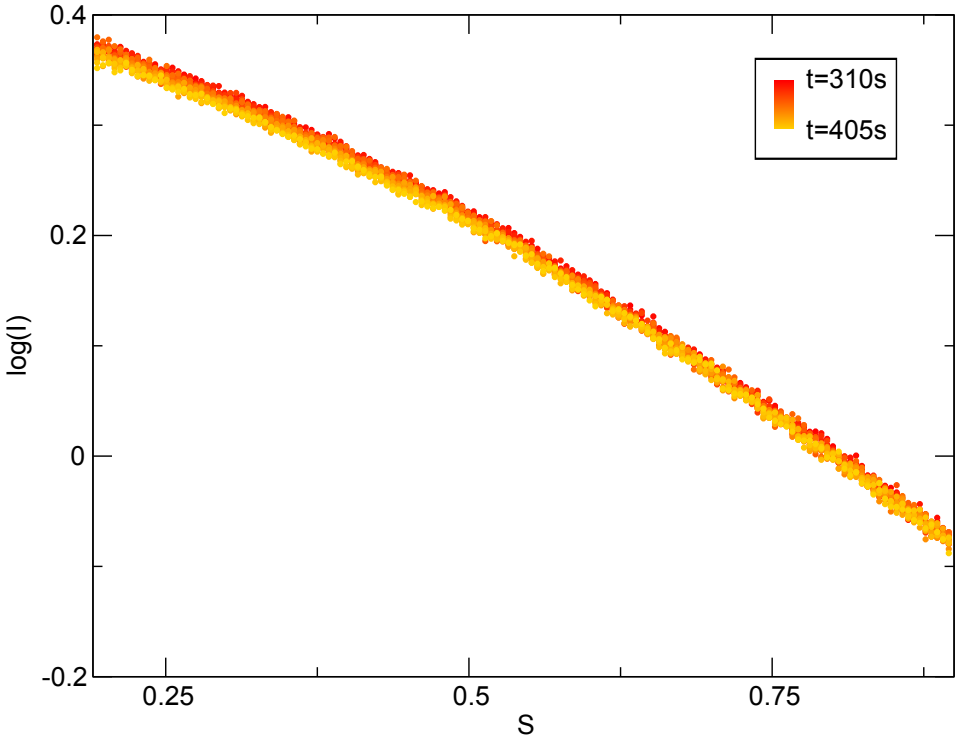


Figure 11.2: SAXS curves evolution, WT B2M, during folding. SAXS curve are obtained as the average of 20 illuminations of 5 seconds each, and subtraction of the closest solvent diffraction curve, using the Bio-SAXS beam-line at the ESRF Synchrotron in Grenoble.



(a)



(b)

Figure 11.3: First (a) and second (b) batch of 20 illuminations obtained during the folding of WT B2M

If standard SAXS curves consist of the average of 10 to 20 illuminations, those illuminations can be considered individually, giving additional time resolution at the expense of the signal to noise ratio. Figure 11.3 displays the 20 illuminations obtained just after the pH jump and the 20 illuminations obtained at t=400s. It is clear that the forward scattering intensity, and thus the average molecular weight, is already decreasing from the beginning of the folding, the first and second SAXS curve of figure 11.2 being the average of these evolutions. Table 11.2 displays the result of the Guinier analysis made on the SAXS curves obtained along the folding, figure 11.4 shows the decay of I(0) and the average molecular weight.

Time (s)	I_0	$R_g(\mathring{A})$	Fidel.	first point used	last point used
100	3,03	23,9	0,99	63	84
360	2,46	22,8	0,993	42	71
860	2,3	22	0,99	41	68
2000	2,18	21,2	0,994	46	73
3600	2,1	20,6	0,942	47	72
7200	2,08	19,9	0,959	47	82
10800	2,09	20,1	0,99	47	74

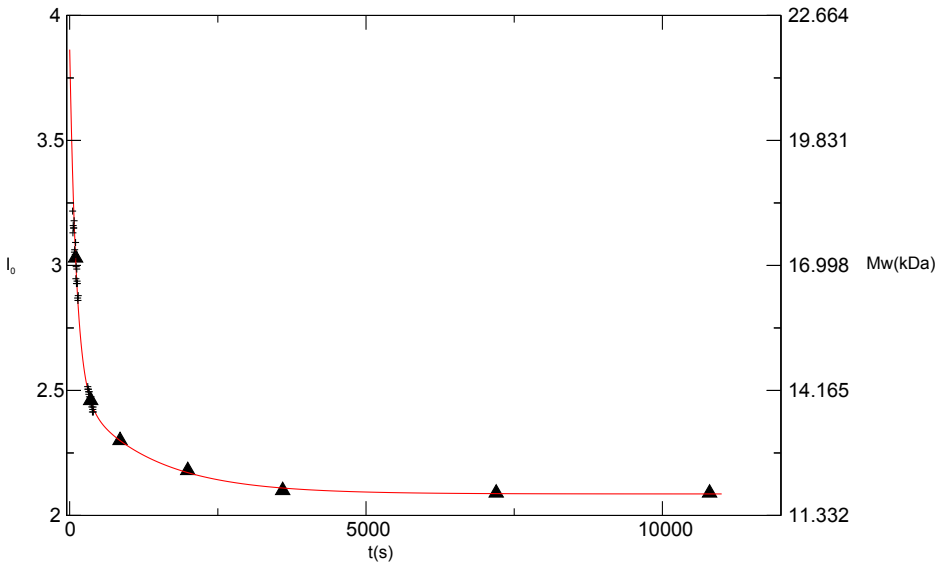
Table 11.2: Guinier analysis along the folding of WT B2M

First observation is that a bi-exponential function is necessary to correctly fit the evolution of the forward scattering intensity. The two rates obtained are $6.8.10^{-3}$ s^{-1} and $7.3.10^{-4}$ s^{-1}. The slow phase of the folding clearly corresponds to what is observed in NMR under similar conditions, while the fast phase is not observed in NMR. This new phase shows the existence of another oligomeric intermediate state, termed I_3. NMR data gives access to the population of the oligomeric intermediate state during the folding. The monomeric intermediate state has the same theoretical molecular weight as the native state monomer, and both populations are also known from NMR data. From these populations and the theoretical molecular weight of the monomer (11.73kDa), the average molecular weight of the oligomeric population at a given time t of the folding can be calculated using equation (11.1), derived from equation (3.3). Results are shown on table 11.3.

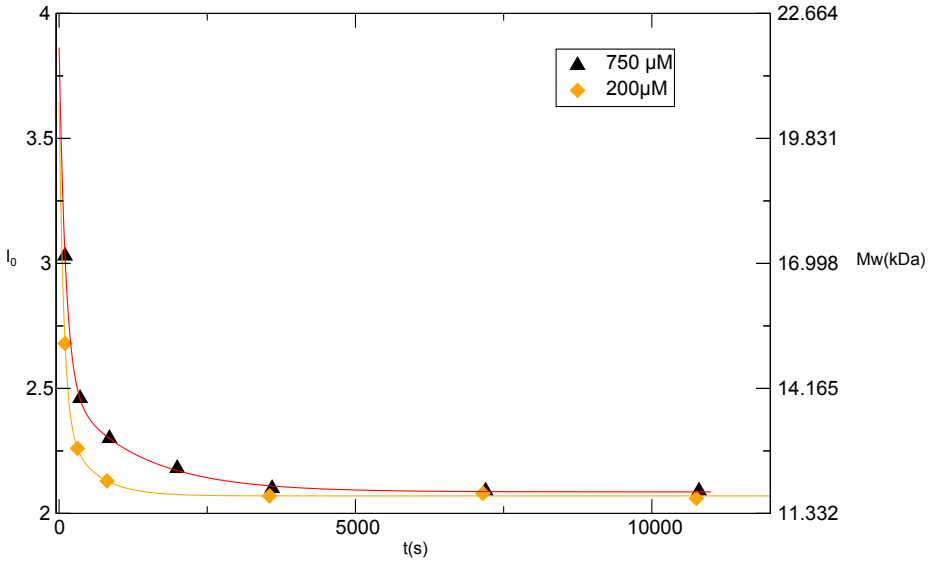
$$\overline{M}_w^{oligo}(t) = \frac{\overline{M}_w^{total}(t) - M_w^{mono} * x^{mono}(t)}{1 - x^{mono}(t)} \tag{11.1}$$

From 6 minutes to 35 minutes after the pH jump, which is the time range in which the fast-phase oligomer I_3 has disappeared and the oligomeric intermediate state I_2 is still significantly populated, the average molecular weight of the oligomeric population is $20kDa$. Considering the underestimation of the molecular weights due to the presence of aggregates in the sample, and concentration imprecision, the low molecular weight oligomer corresponding to the slow-phase folding I_2 seems to be a dimer.

This experiment was repeated with a lower protein concentration (200µM final) as



(a)



(b)

Figure 11.4: Evolutions (black triangles or orange diamonds) and fits (solid line, red or orange) of I(0) along the folding of B2M at a concentration of 750µM (a) and 200µM (b). On (a), + correspond to the values obtained from individual illuminations. The right y axis gives the average molecular weight corresponding to the I(0) value on the left y axis, evaluated using a comparison to the BSA.

Time(s)	I_0	$\overline{M}_w^{total}$	x^{mono}	$\overline{M}_w^{oligo}$
100	3,03	17,1	0,75	33,0
360	2,46	13,9	0,79	22,0
860	2,3	13,0	0,86	20,2
2000	2,18	12,3	0,94	20,3
3600	2,1	11,9	0,98	-
7200	2,08	11,8	1	-
10800	2,09	11,8	1	-

Table 11.3: Average molecular weight of the oligomeric fraction. Monomeric fraction is the I+N fraction measured in NMR at T=20°C.

shown on figure 11.4. First remark is that the average molecular weight of the sample during the folding is lower than at high concentration. This means either that the stoichiometries of the oligomeric species are lower or that they are less populated. NMR data, along with common thermodynamic considerations, suggests that indeed the population of oligomeric species decreases with concentration. If the population measured in NMR at 200µM is taken into account, between 6 and 35 minutes, the molecular weight is of $20kDa$, the same as at high concentration, consistent with the low molecular weight oligomer I_2 being a dimer.

Second remark is that the folding fast phase is also less important than at high concentration, leading to the same hypothesis as for the lower molecular weight oligomer, either a lower stoichiometry or population. As for the high concentration experiment, no insights on the population of this state can be obtained from the NMR data, it is therefore difficult to evaluate the molecular weight of this oligomer. But since it seems to be undetectable in NMR, the population is likely to be small compared to that of the lower molecular weight oligomer. To have such a visible impact in SAXS, I_3 is therefore probably much bigger than a dimer.

11.2 Infrared Spectroscopy

As shown in section 3.4, Attenuated total reflectance (ATR) fourier-transform infrared spectroscopy (ft-IR) can be used to detect secondary structures, especially when considering the Amide I absorption band. This technique may yield information on the secondary structure of the intermediate states.

The ATR ft-IR spectrometer requires the deposition of a drop of the liquid sample on its germanium crystal, which implies the use of 10µL for each measurement, to avoid a bias due to the drying of the drop. In these ATR ft-IR experiments, folding is initiated using by-hand mixing, and the deposition of the sample drop inside the spectrometer implies a dead time of about 30 seconds. Thanks to the properties of ATR spectrometers, buffers that are not deuterated can be used, offering the possi-

bility to use the same buffers as in other experiments. These experiments were lead at room temperature, and a concentration of 750µM final.

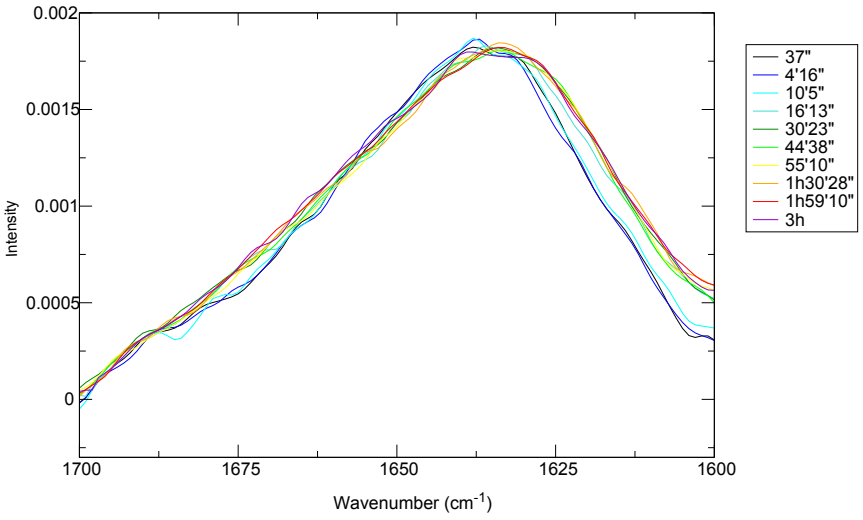


Figure 11.5: Evolution of the infrared spectrum of B2M along the folding. Spectra are obtained using ATR ft-IR. The folding is initiated in a reactor, and samples are analyzed along the folding. Experiments were lead at room temperature, and a concentration of 750µM final.

The IR signal corresponding to the amide I band is plotted on figure 11.5, along with its evolution during the folding. As stated in section 3.4, the amide I band can be used to estimate the secondary structure content of a protein, or its changes along time. Since it is composed of the vibration of all the C=O groups of the backbone of the protein, the obtained information is rather a global trend than a precise structural information. Here, along the folding, intensity is increasing in two bands, around 1620cm-1 and 1675cm-1, which are specific of beta sheets, and decreasing around 1640cm-1, which can be specific of either unstructured regions or alpha helices. Since the context of the observation is the folding of B2M, this should be interpreted as the disappearance of unstructured regions, along with an increase in the beta sheet content. This means that the main event observed in infrared spectroscopy is the transition between the monomeric intermediate state and the native state, and not the interconversion between the monomeric and the oligomeric intermediates. This is not such a surprise, for two reasons. First, the population change between I_1 and I_2 represents only about 20% of the total signal at this temperature and concentration, while the population change between I_1 and the native state represents about 70%. The disappearance of intermolecular beta sheets with the disappearance of the I_2

state is therefore compensated by the folding to the native state, with the formation of intramolecular beta sheets.

The use of a single set of experimental conditions does not allow to conclude on the existence or absence of intermolecular beta sheets at the beginning of the folding, because undistinguishable intramolecular beta sheets form at the same time. But if concentration could be lowered, as much as I_2 would only be negligibly populated, the relative change in the beta sheet content would be increased, proving the compensation at high concentration by the disappearance of intermolecular beta sheets. Yet at 9mg/mL, we are already at the limit of detection of the spectrometer that was used, which means that further work of optimisation of the experimental conditions would be necessary to obtain that kind of information.

11.3 UV-fluorescence spectroscopy with the use of Thioflavin-T

As seen in section 3.2, Thioflavin-T is a fluorescent reagent specific of amyloid structures, binding inside long stabilized beta-sheets. It does not require the formation of full-length fibrils, but also binds to large oligomers that have an amyloid-like structure. Following the folding of B2M in presence of Thioflavin-T may be a way to characterize the oligomeric intermediate states.

The experiment was made using a signal stabilized through the measurement of the variations of the incident beam. The experiment was run at room temperature. Buffer conditions are similar to those used in SAXS or IR, and the final concentration is also 750µM. Thioflavin-T is mixed in the refolding buffer, which is added to the unfolded protein to induce the pH jump. The deadtime corresponds to the time needed to mix the refolding buffer and the unfolded protein inside the measurement cell, to put this cell inside the spectrometer, and to start the measurement, and is evaluated to circa 40 seconds.

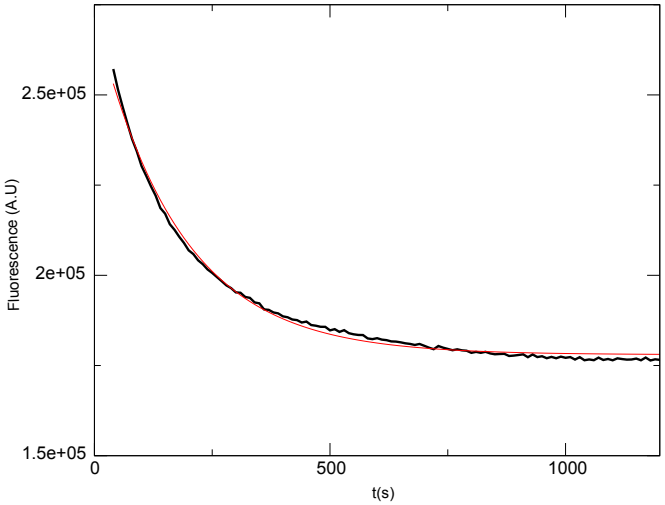


Figure 11.6: Folding followed with UV fluorescence, in presence of Thioflavin-T (black), and corresponding fit (red)

Results (figure 11.6) show a decrease of the Thioflavin-T fluorescence at the very beginning of the folding with a kinetic rate of $5.6.10^{-3}\ s^{-1}$, and for the rest of the folding, a constant signal. This means that Thioflavin-T binds one of the intermediate states, which disappears within 5 minutes. It corresponds to the higher molecular weight oligomeric intermediate state I_3 detected in SAXS, which decays with a rate constant of the same order of magnitude.

In conclusion, following the folding of B2M in presence of Thioflavin-T allows us to further characterize the oligomeric intermediate states. Indeed, at least two states, or ensemble of states, can be distinguished. One is a long-lived, low molecular weight intermediate state, that does not bind Thioflavin-T, either because is does not adopt an amyloid like structure, or because of its too low molecular weight, termed I_2. The other is a short-lived, higher molecular weight intermediate state, which binds Thioflavin-T, and as such adopts an amyloid like structure, termed I_3.

11.4 Back to NMR, detection of a short lived intermediate state

A full investigation of the NMR spectra acquired during the folding of B2M, based on the findings in SAXS and UV-fluorescence, reveals the existence of three $^1H - ^{15}N$ correlation peaks that disappear a few minutes after the pH jump, as shown on figure

11.7 and 11.8. This finding is consistent with the hypothesis of a short lived intermediate state, as seen in SAXS and UV-fluorescence. These peaks may correspond to a flexible part of the protein (the C-terminus for example) in the oligomeric state, allowing their detection while the rest of the protein remains NMR-invisible due to the slow tumbling of these oligomers in solution.

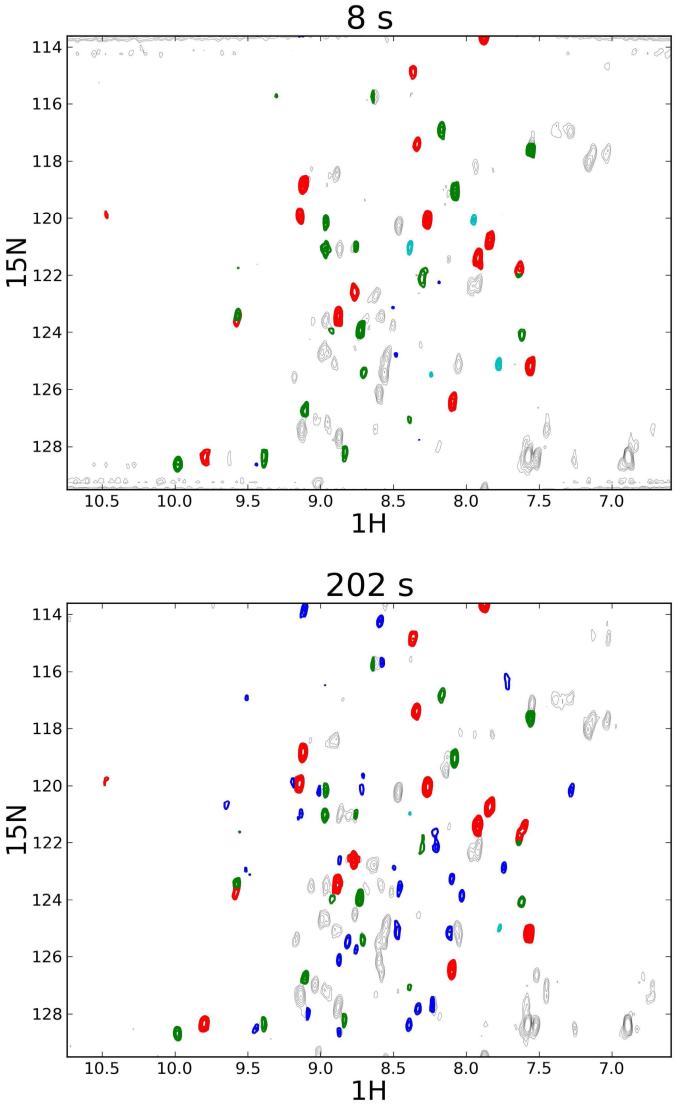


Figure 11.7: Spectra of WT B2M just after the pH jump and 200 seconds later, obtained using FTA-SOFAST. Different classes of peaks are highlighted: I_1 in green, N in dark blue, I_3 in light blue, $I + N$ in red, showing the existence of a species that disappears within a few minutes.

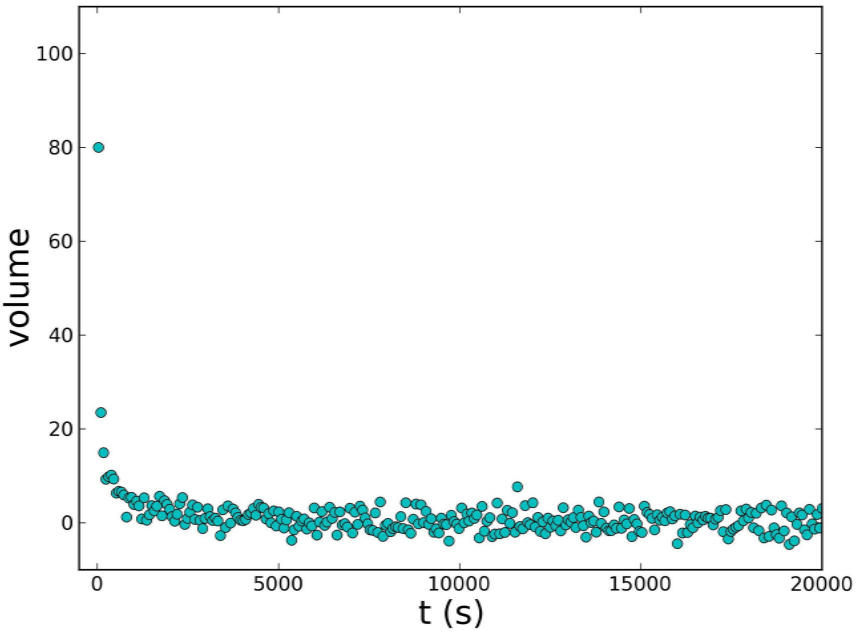


Figure 11.8: Evolution of the volume (normalized to fitted initial value) of the peaks attributed to I_3 during the folding of WT B2M

11.5 Conclusions

From NMR data, the NMR-invisible state I_2 could be assigned to an oligomeric state. From the SAXS and UV-fluorescence data, this oligomeric state has to be split into two states, one long lived low molecular weight oligomer, probably a dimer which will still be named I_2, or ensemble of long lived low molecular weight oligomers, and a short lived higher molecular weight oligomer or ensemble of oligomers, I_3.

Clearly the size of the higher molecular weight oligomer(s) I_3 explains its almost NMR-invisibility. But a dimer, or an ensemble of low molecular weight oligomers (dimer, trimer, tetramer) should be visible on the NMR spectra. These oligomeric states I_2 are in equilibrium with I_1. Three hypotheses were drawn in chapter 10.3.2 to account for the data:

- $I_1$and I_2 are in slow exchange. I_1 is monomeric and NMR-visible, while I_2 is oligomeric and invisible. This means that I_2 is a large oligomer, so that the peaks are broaden by its relaxation properties due to a large tumbling correlation time.
- $I_1$and I_2 are in fast exchange. I_1 is monomeric, while I_2 is a low molecular weight oligomer. Therefore the intermediate state spectrum that we see in NMR is a weighted average of the I_1 and I_2 states spectra. But both the linewidth and the

T_1 relaxation parameter should be different between this averaged intermediate state and the native state.

- $I_1$and I_2 are in slow exchange. I_1 is monomeric, while I_2 is a low molecular weight oligomer that should be NMR-visible. I_1 and I_2 spectra exactly overlap, so that only one spectrum is visible. But again both the linewidth and the T_1 relaxation parameter should be different between this averaged intermediate state and the native state.

From the SAXS data, and the fact that I_2 is probably a dimer, the first hypothesis has to be discarded. This means that the criteria of the linewidth comparison with the N state and the longer recycling delay experiment are not precise enough to give reliable information on the equilibrium. The two remaining hypotheses can not be discriminated based on the data available, therefore further work of characterization is to be done, using for example spin relaxation measurements in real time NMR.

The use of SAXS, NMR and UV-fluo as complementary methods provides consistent data, allowing a fine characterization of the B2M folding process. A large amyloid-like oligomeric state is lowly populated at the very beginning of the folding, while a long-lived low molecular weight oligomer is present in important proportions until the protein is properly folded. This discovery sheds a new light on the folding of B2M, providing bridges between the folding and the aggregation processes. The oligomeric states transiently populated during folding maybe the same as the intermediate states on the pathway to the formation of amyloid fibrils. Putative folding model is shown on figure 11.9, where the involvement of I_3 remains unclear.

The folding data, both NMR and SAXS, can be fitted altogether using this model, as shown in figure 11.10, with good agreement, parameters being used for the fit are the equilibrium constant between I_1 and I_2, the isomerization rate of the Proline 32, and the initial populations of N at the different concentrations. Obtained values are similar to those previously observed: $8.2.10^{-4}s^{-1}$ for the isomerization rate, and around 10% for the N-state burst phases respectively. One new element from this global fit is the equilibrium constant of the $I_1 - I_2$ equilibrium: $K_{I_1-I_2} = 552$. Of course the fit is not perfect: for example at 750µM, the I_1 to N conversion seems to be faster than what is expected using the model. This can be explained: the model used implies that the I_1 to N conversion rate is proportional to the initial population of I_1, which depends on the concentration. Since such a dependence is not observed in the NMR data, the folding mechanism is probably more complex than this model, maybe with a dimeric native state in equilibrium with the monomeric native state, and a new parallel folding pathway from I_2 to this "N_2" state.

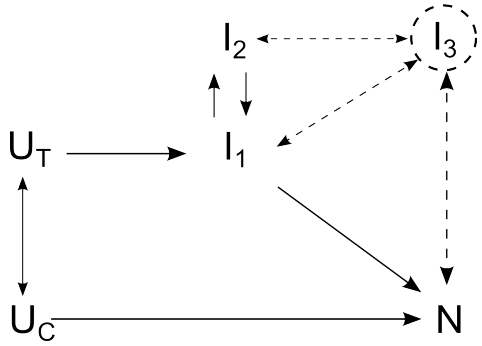
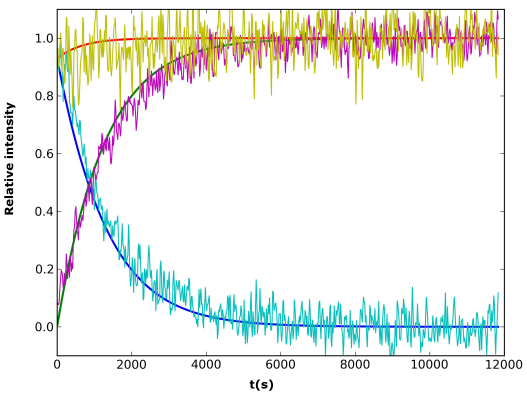
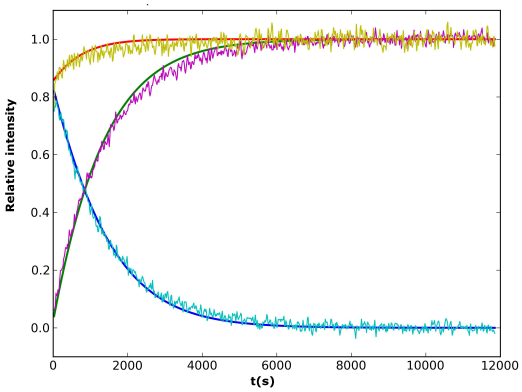


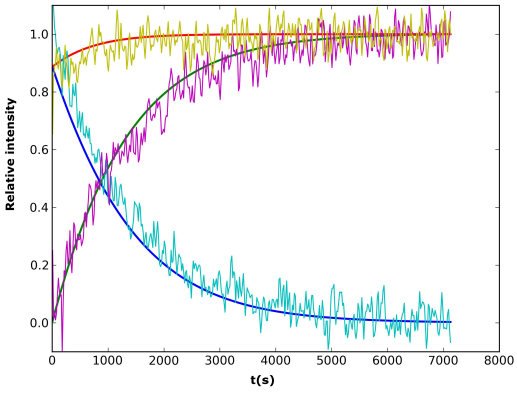
Figure 11.9: Putative model for the folding of B2M. The location of I_3 in this mechanism is still unknown.



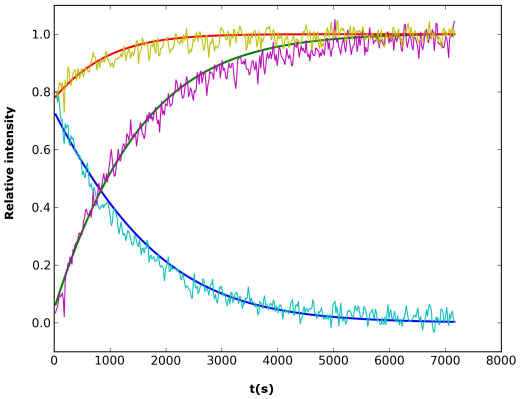
(a) C=0.05mM



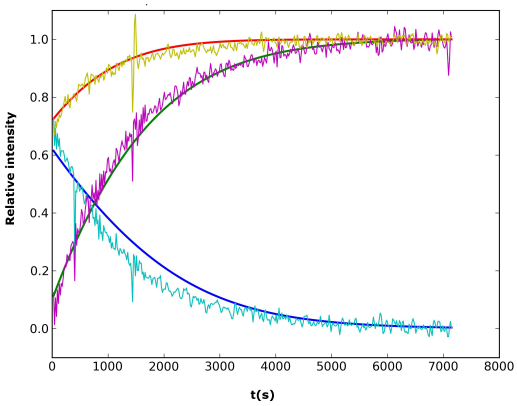
(b) C=0.15mM



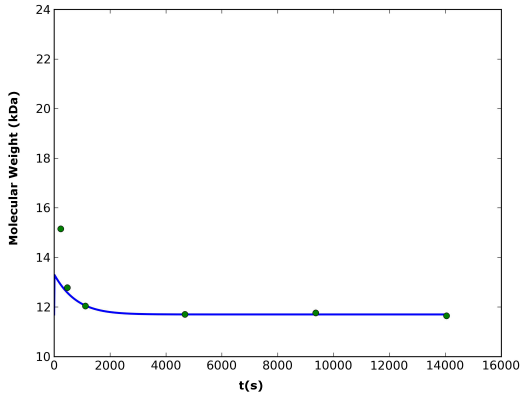
(c) C=0.188mM



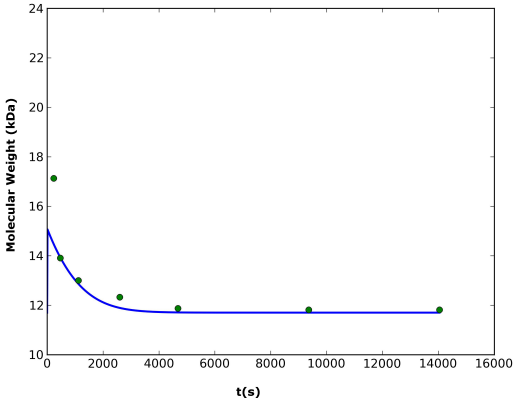
(d) C=0.375mM



(e) C=0.750mM



(f) C=0.200mM



(g) C=0.750mM

Figure 11.10: Experimental data of the folding of B2M at 20°C fitted using the model presented on figure 11.9, with NMR data (a)-(e) (the N class in green, I class in blue, I+N class in red, intensities of peaks summed inside each class, and normalize to initial or final value) and SAXS data (f) and (g). Fits are represented by solid lines.

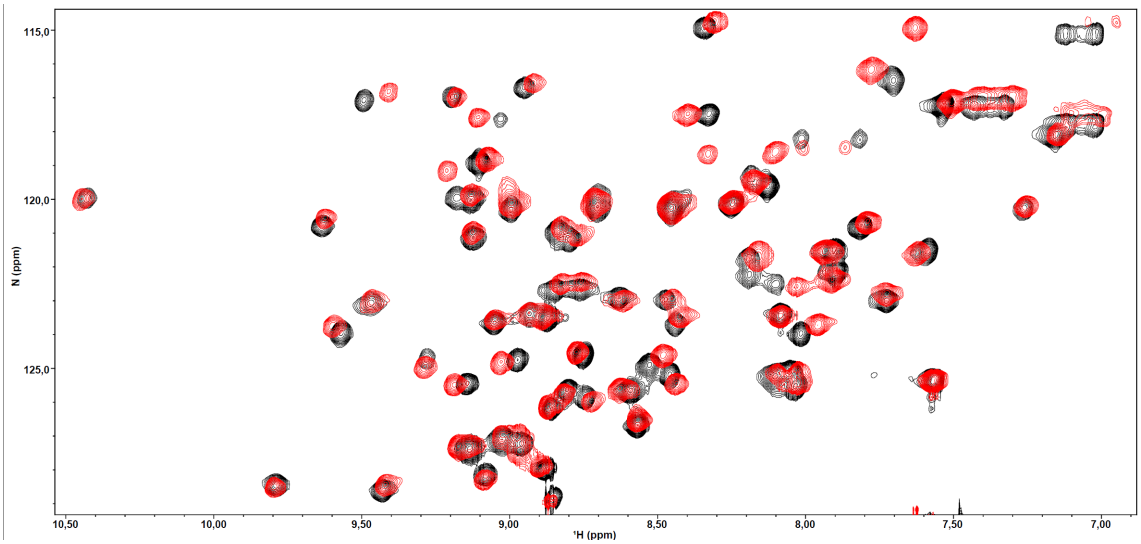
Chapter 12

Single point mutants as in-equilibrium models?

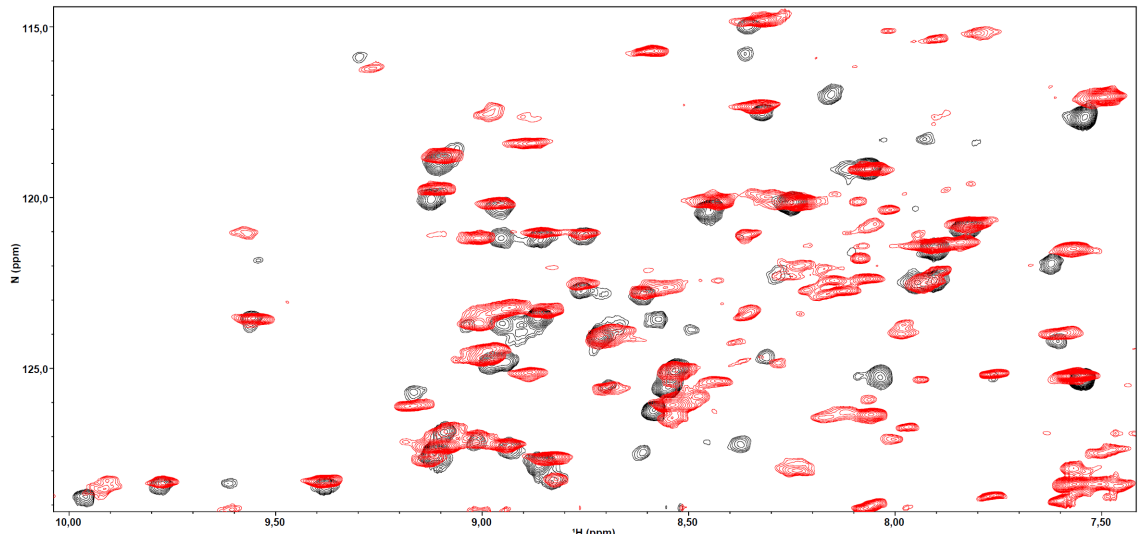
Creating protein mutants as in-equilibrium models of short-lived states of that protein is a method of choice in structural and functional biology. This method has been applied to B2M, as was seen in chapter 5, for example with the use of the P32A B2M mutant. P32A has as main characteristic a trans peptide bond between residues 31 and 32, and as such is presented as mimicking the trans conformation of the proline 32 either during folding or amyloid fibril formation. Data obtained on I_t offers the possibility to test the capacity of P32A to be a good in-equilibrium model for this transient state. Does it populate oligomeric states, with both low and higher molecular weight? Are the characteristics of the equilibrium between these states, if they can be detected, similar to those observed for I_t, ie model the I_1-I_2 equilibrium and the presence of I_3?

Another single point mutant that may be used as a model is W60G. It has been shown to be less aggregation-prone (section 5.2) both in equilibrium and during folding, and to have a simpler folding mechanism (section 8.3). It may be an in-equilibrium model of a purely monomeric B2M native state.

As shown on figure 12.1, the NMR spectra of P32A and W60G are very similar to those of the WT I state and N state, respectively. This is a first "footprint" validation of these single point mutants as in-equilibrium models of different states of the WT B2M.



(a)



(b)

Figure 12.1: Spectral comparison of W60G (red) and the N state of the WT protein (black) (a) and of P32A (red) and the I state of the WT protein (black) (b). In (b), the different linewidth in the indirect dimension is due to different experimental setups.

12.1 Evaluating the average molecular weight using SAXS

12.1.1 Comparison of the in equilibrium SAXS curves of WT, W60G and P32A B2M

As seen in section 3.5, SAXS can be used to measure the average molecular weight in a sample. If oligomers are present, the average molecular weight will be higher than that of the monomer. Figure 12.2 displays the SAXS curves of WT B2M and the two mutants P32A and W60G.

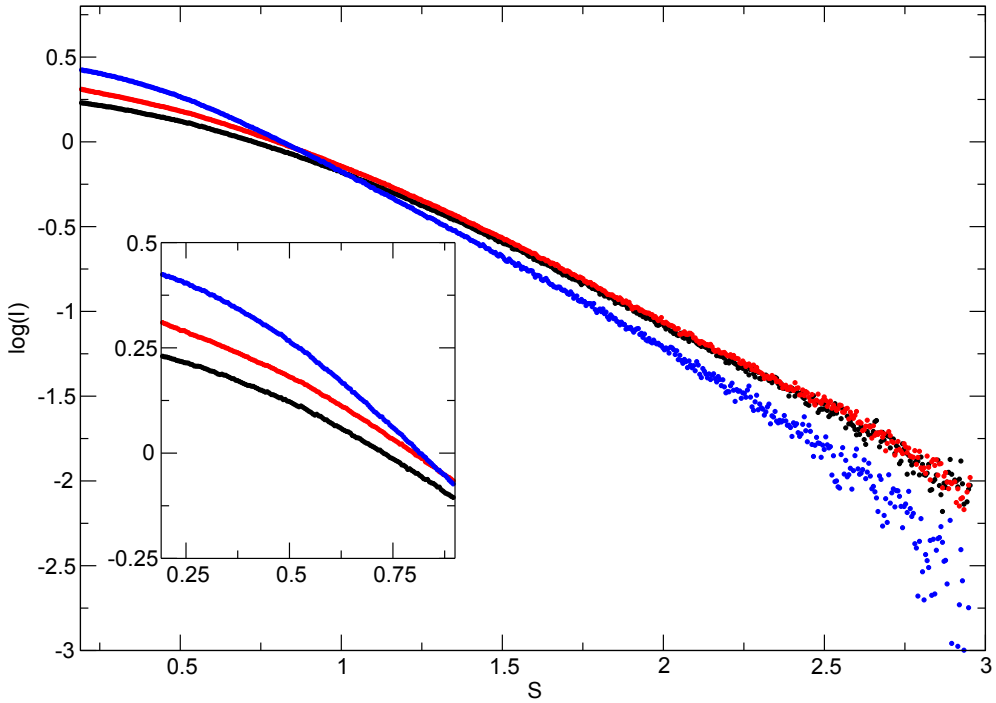


Figure 12.2: SAXS curve of P32A 5.36mg/mL (blue), WT 6.75mg/mL (red), W60G 6.35mg/mL (black), obtained using the average of 10 illuminations of 10 seconds each.

It is evident from this data that the average molecular weight of P32A is higher than that of WT, that is higher than that of W60G (see also table 12.1). This means that the oligomeric fraction of P32A is much higher than the other two species, but also that the WT B2M is not monomeric in these conditions. With an average molecular weight lower than that of the WT protein, W60G appears as a model for a monomeric state of B2M. The fact that the measured molecular weight is lower than the theoretical one is due to the presence of large aggregates in solution, leading to an under estimation of the measured molecular weight. This data for the WT B2M is consistent with the existence of an oligomeric native state in equilibrium conditions.

	Concentration (mg/mL)	Io	Rg (Å)	Average Molecular Weight (kDa)
P32A	5,36	2,83	23,3	20,1
WT	6,75	2,09	20,1	11,8
W60G	6,35	1,78	19,4	10,06

Table 12.1: Guinier analysis of P32A, WT and W60G B2M in equilibrium

12.1.2 Further characterization of the oligomerization of P32A

As seen before, the average molecular weight of P32A in equilibrium is 20.1 kDa at 400µM, clearly higher than that of WT B2M (12kDa). This means that in equilibrium P32A populates oligomeric states, just as WT B2M does during folding. This does not give insights on the stoichiometry of this oligomeric state, or even if several oligomeric states are populated. Equation (12.3) shows that, in the case of a two state equilibrium (12.1), the stoichiometry can be linked to the variation of the population of the oligomeric state and the variation in the normalized forward scattering intensity. Practically, considering the extrapolation to infinite dilution, where the population of the oligomer is zero, and the average molecular weight is the molecular weight of the monomer, stoichiometry is linked to the population of the oligomer at a given concentration (12.5).

$$n.A \rightleftharpoons A_n \tag{12.1}$$

$$\frac{I_0}{C_{tot}} = \frac{I_{0BSA}}{M_{BSA}}[x.M_A + (1-x).M_{A_n}] \tag{12.2}$$

$$\frac{I_0}{C_{tot}} = \frac{I_{0BSA}.M_A}{M_{BSA}}.[x + (1-x).n]$$

$$\frac{I_0}{C_{tot}} = \frac{I_{0BSA}.M_A}{M_{BSA}}.[n + (1-n).x] \tag{12.3}$$

$$\alpha_{I_0} = \frac{d(\frac{I_0}{C_{tot}})}{dC_{tot}} = \frac{I_{0_{BSA}}.M_A}{M_{BSA}}.(1-n).\frac{dx}{dC_{tot}} \tag{12.4}$$

$$\frac{\triangle C_{tot}}{\triangle x}.\alpha_{I_0}.\frac{M_{BSA}}{M_A.I_{0_{BSA}}} = 1-n \tag{12.5}$$

Based on this, the SAXS curves of successive dilutions of P32A (figure 12.3) can be used to extrapolate the molecular weight of ideal monomeric P32A, and as such the fractions of oligomers that would be needed to assess for the SAXS data. This is still not sufficient to definitely characterize the stoichiometry. Equation 12.5 can be

used to calculate the stoichiometry n in function of the oligomeric fraction at a given concentration, with α_{I_o} obtained from the linear regression shown on figure 12.4. As shown on figure 12.5, if P32A oligomeric fraction is similar to that observed during the folding, then it is likely to populate a range of oligomers going from dimers to tetramers.

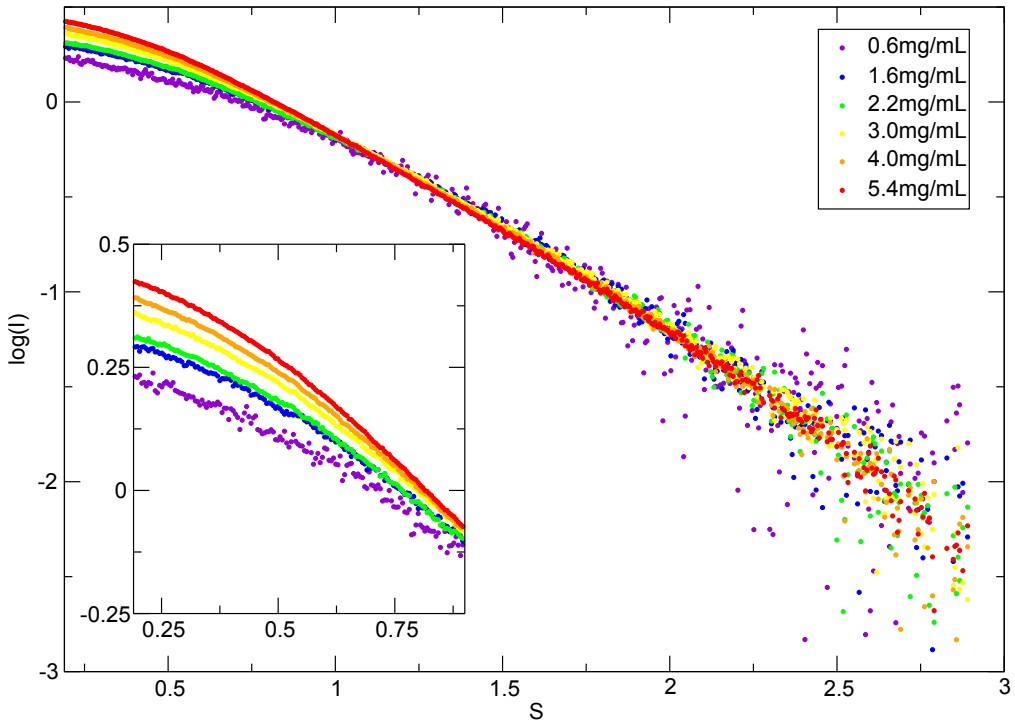


Figure 12.3: SAXS curves of successive dilutions of P32A, obtained using the average of 10 illuminations of 10 seconds each.

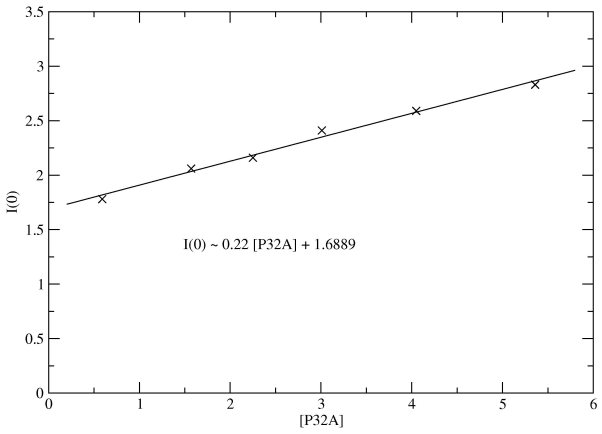


Figure 12.4: Variation of the normalized forward scattering intensity against the total concentration of P32A. Equation corresponding to a linear regression is shown on graph.

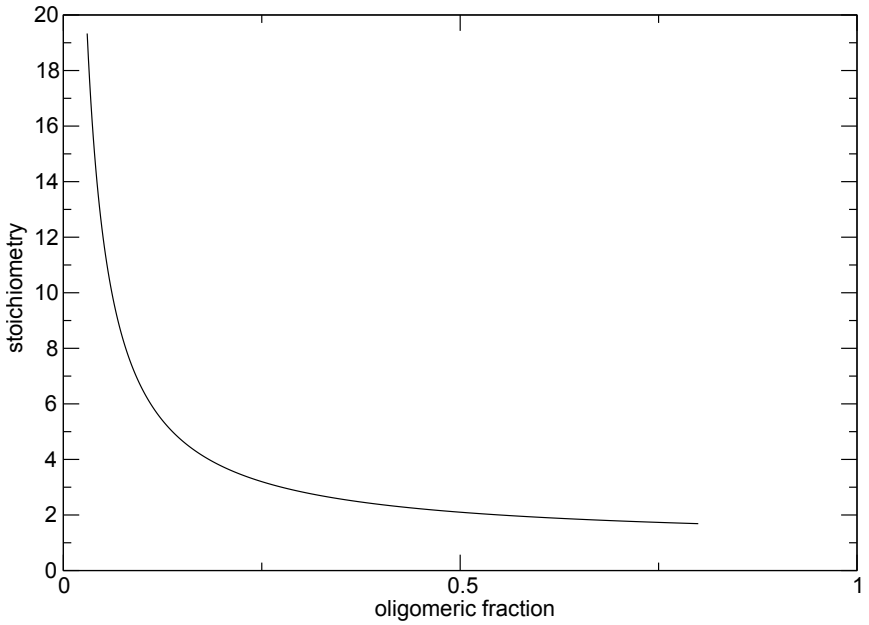


Figure 12.5: Calculated stoichiometry of the oligomeric state of P32A in function of the oligomeric fraction at a total P32A concentration of 6.5mg/mL

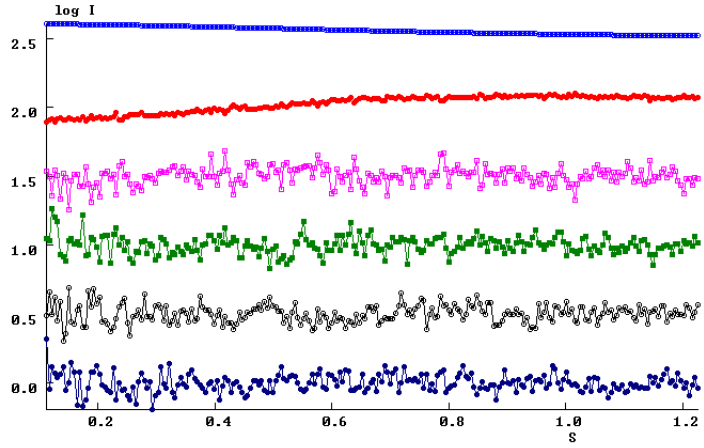


Figure 12.6: Singular Value Decomposition of the dilution of P32A, as calculated using the SVD-plot utility of the PRIMUS software.

In order to check the two-state model hypothesis, singular value decomposition can be performed on the dilution data. Singular value decomposition is a statistical tool that can detect the minimal number of components (eigenvectors) that are needed to assess for the global signal of a mixture, here the number of states populated by P32A. The eigenvectors corresponding to the dilution data are displayed on figure 12.6. Even though eigenvectors from the third from the top (pink trace and below) seem to contain essentially noise components, the statistics given by the singular value decomposition state that four states are populated, one monomeric, and three oligomeric. From this statistical analysis, P32A seems to be populating an ensemble of small oligomers, from dimer to tetramer.

But again, if P32A populates a dimeric state in equilibrium, shouldn't it be visible in NMR? If the dimeric and monomeric states are in slow exchange, hypotheses to explain this invisibility are either that the dimer is highly dynamic, or that it is itself in fast exchange with the higher molecular weights oligomers: trimer and tetramer. If the two states are in fast exchange, then the peaks that appear on the spectrum are the weighted average of those of the two states. In that case of fast exchange, the oligomerization of P32A should be detectable also using NMR relaxation measurement experiments.

12.2 NMR ^{15}N relaxation experiments

^{15}N relaxation experiments consist in the measurement of the T_1 (longitudinal) and T_2 (transverse) relaxation times of all amide nuclei along the protein backbone.

T_1 and T_2 measurements along the sequence of P32A are shown on figure 12.7. The T_1/T_2 ratio, linked to the τ_c of the protein by equation 7.2 where ν_N is the resonance frequency of Nitrogen at the considered magnetic field, is shown on figure 12.8. It can be compared to the T_1/T_2 ratio of another mutant of B2M, W60G (also on figure 12.8), that is considered a monomeric state model of the protein. The T_1/T_2 ratio of P32A is significantly higher than those of W60G, consistent with P32A populating a higher τ_c state, in this case an oligomeric state. Equation 12.6 shows the correlation for monomeric proteins between the τ_c and the molecular weight [141].

$$\tau_c(ns) \approx 0.6 M_W(kDa) \tag{12.6}$$

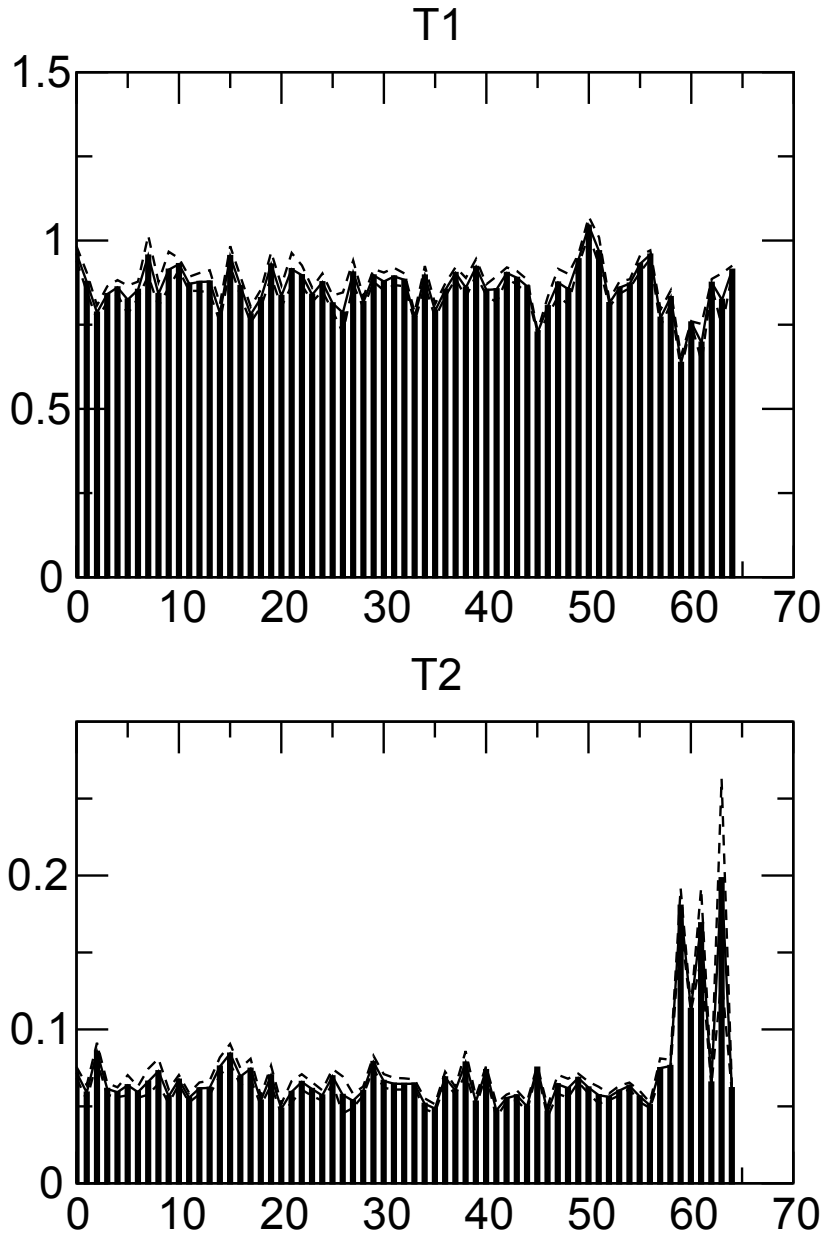


Figure 12.7: ^{15}N T_1 and T_2 (in seconds) of P32A at 750µM with error bars, versus arbitrary residue numbers

The oligomerization of P32A can further be characterized using the same experimental set-up as in SAXS: dilution. Figure 12.8 shows the T_1/T_2 ratios of P32A and W60G at 750µM and 180µM. While the T_1/T_2 ratios of W60G are mostly insensitive to dilution, P32A displays a sharp decrease of its T_1/T_2 ratios, and consequently its τ_c, to reach at low concentration values equivalent to those of W60G. P32A is close to monomeric at a concentration of 180µM.

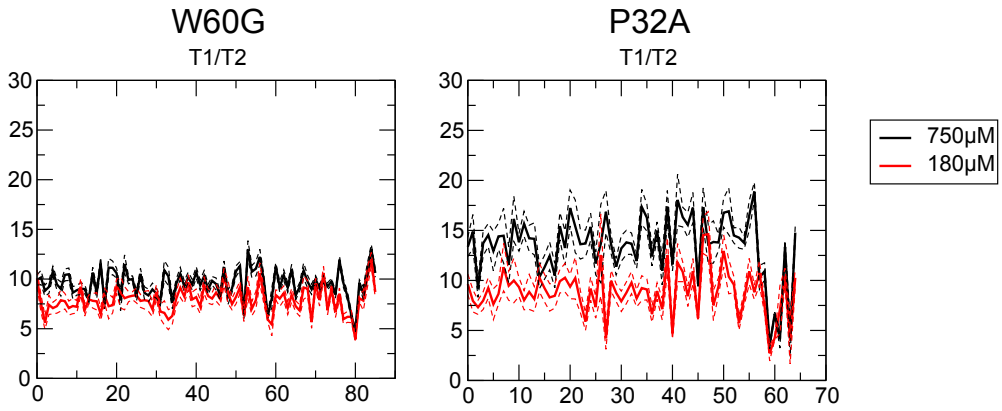


Figure 12.8: ^{15}N T_1/T_2 ratios of P32A and W60G at 750µM (black) and 180µM (red), against arbitrary residue numbers.

12.3 Conclusions

To sum up the findings of this chapter,

- W60G is a good in-equilibrium model of the monomeric native state of B2M

W60G has a lower average molecular weight than the WT protein, in equilibrium, as seen in SAXS, which means that W60G is less prone to populate oligomeric states under these conditions. As shown by relaxation NMR, W60G is almost insensitive to dilution, which implies that oligomeric states are not populated in equilibrium, even at high concentrations. From those two results, W60G can be considered as a monomeric model of B2M native state.

- P32A may be an in-equilibrium model of the I_1-I_2 equilibrium

P32A has a NMR spectrum that is very similar to that of I_1-I_2, forms low molecular weight oligomers, probably a ladder from dimer to tetramer. These are characteristics that are common to P32A and I_1-I_2. But on the other hand, P32A doesn't form higher molecular weight oligomers, like I_3. Relaxation results would have to be compared to measurements made in real time on I_1-I_2, but are compatible with proposed timescales of this equilibrium.

Part V

Toward NMR structural characterization of short-lived folding intermediates

Structural characterization of proteins using NMR is highly demanding in experimental time. To assign a protein, and link the data to specific residues, a set of 3D experiments are needed, corresponding usually to a few hours of measurement time for each spectrum. Structural data also demands long experimental times, for example when using $^1H -^1 H$ NOE or residual dipolar couplings (RDC).

That kind of standard strategies to obtain structural data is not suited for the study of short-lived species such as folding intermediates. Therefore we have developed a two step strategy: first optimizing the signal to noise ratio that can be obtained during the lifetime of the sample, either by increasing this lifetime, or by optimizing the experimental conditions; second using techniques that are less experimental time demanding. This part shows this two step strategy and draws perspectives on the development of methods for the fast acquisition of structural data.

Chapter 13

B2M W60G as a test case

13.1 Cumulative discrete data acquisition vs continuous data acquisition

When studying the kinetics of the folding of B2M, the balance has to be made between time resolution and signal to noise ratio. As peaks are classified in classes, in which the signal is summed, the signal to noise ratio needed is lower than if we were studying individual peaks. This is why the experimental time required for the acquisition of one spectrum could be reduced to about 15 seconds.

But obtaining residue resolved structural information demands the use of experimental time consuming 3D NMR methods, of the order of hours. A clear difficulty when the subject of study is a transient, short-lived state. A first step is therefore to optimize the signal that can be obtained during the lifetime of the protein state. For B2M, the intermediate state has a lifetime of about half an hour at 15°C.

A single standard 3D experiment run during the whole lifetime of the intermediate would be an experiment during which the sample is evolving, with population decreasing for the intermediate state, which would imply line broadening. Two solutions can be considered to assess for this line broadening, either accumulating successive acquisitions, or using apodization to correct for a continuous acquisition.

13.1.1 Cumulative discrete data acquisition

The idea is to run an experiment on a time frame during which the sample can be considered unchanged, ie small compared to the timescale of the folding. The experiment can then be repeated over successive time frames, and the signal summed to obtain one higher signal to noise ratio spectrum. This is the approach that has

been used in the previous paragraphs of our study. For the reconstruction of pure I-state NMR spectra, two points need further consideration.

First, after a certain point in the folding, the population of the intermediate state is so low that adding more NMR spectra would only result in the addition of noise, and as such a reduction of the overal signal to noise ratio. As shown on figure 13.1, after 25 minutes, the addition of more spectra implies a decrease of the signal to noise ratio of intermediate state peaks, so this limit can be considered the NMR effective lifetime of this state under a given set of experimental conditions.

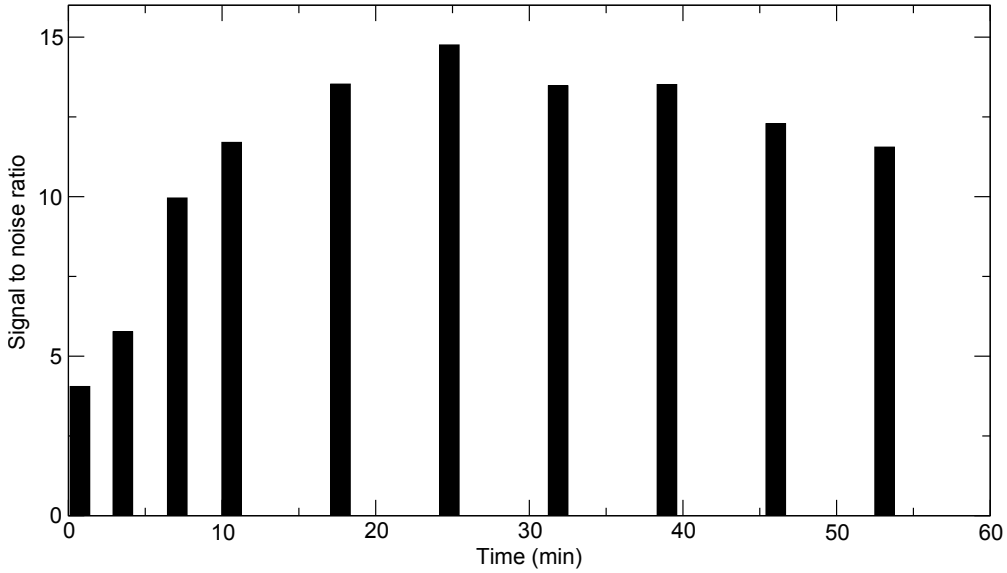
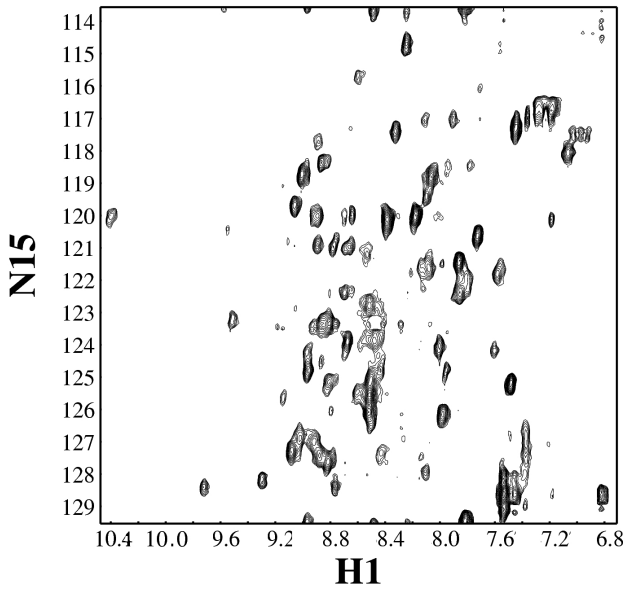


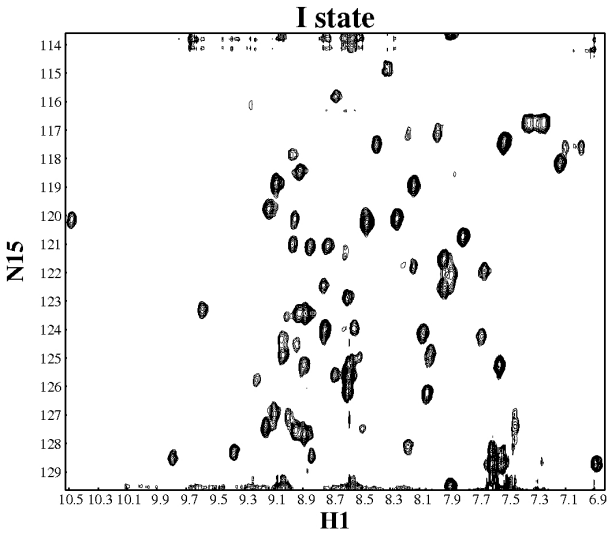
Figure 13.1: Average Signal to Noise ratio of the reconstructed I-state spectrum along the summation of successive spectra during the folding of the protein (HEPES buffer, 750µM, 25°C)

Second is the fact that the protein molecules leaving the intermediate state do not disappear, they fold into the native state. As was shown earlier, the spectrum of the native state partially overlaps with that of the intermediate state. Adding the signal from successive time frames during the folding implies that a growing signal from the native state is added, a signal that can be monitored through the NMR signals of specific, isolated, native state peaks. Simply adding spectra of successive time frames does not result in an intermediate state spectrum, but in a I+N mixed spectrum. But since there are isolated peaks specific of the native state, the "amount" of native signal in this I+N spectrum can be measured. The subsequent subtraction of a highly defined native-state spectrum, normalized with the measured native signal in the I+N spectrum, results in a spectrum of only the intermediate state, with an increase of the signal to noise ratio by more than 3 fold with respect to a single NMR spectrum recorded at the beginning of the folding process, as shown on figure 13.2.

This method allows the use of conventional NMR methods, in terms of pulse sequence



(a)



(b)

Figure 13.2: First spectrum after pH jump (a) and reconstructed I state spectrum (b), using NMR data obtained during the folding using the SOFAST pulse scheme

or processing, yet it requires that the experimental time of one acquisition remains small relatively to the evolution timescale of the studied phenomenon. For B2M folding, this allows the use of fast pulsing 2D methods (such as SOFAST or BEST).

13.1.2 Continuous data acquisition

During the acquisition of the real-time spectrum, the I-state converts into the N-state, therefore the resonances of both the I- and the N- states are sampled. The reconstruction of the I-state spectrum is obtained by a proper combination of the real-time spectrum and the steady-state spectrum. It is important to keep in mind that in the real-time spectrum, the three dimensions differ by the times effectively needed for the frequency sampling (Δt_{real} : time separating two consecutive sampling points):

- for the 1st dimension, that corresponds to direct 1H frequency sampling, consecutive sampling points are separated by the NMR dwell time, and the sampling is completed within 50 to 100 ms;
- for the 2nd dimension, i.e. indirect sampling of the first heteronuclear dimension (either 13C or 15N) in the inner loop, the separation of two sampling points corresponds to the time required to record a 1D spectrum (including phase cycling), typically 200 – 500 ms using BEST-TROSY techniques, yielding a effective acquisition time along this dimension on the order of less than 1 minute;
- for the 3rd dimension, i.e. indirect sampling of the second heteronuclear dimension (either 13C or 15N) in the outer loop, the effective acquisition time is equal to the full duration of the experiment, i.e. around 40 minutes.

The difference in effective sampling times implies that the effect of the folding kinetics on the peak shapes can be safely neglected during both t_1 and t_2, while it will surely affect the t_3 dimension. Hence, because of the convolution with the respective kinetic folding profiles in the third spectral dimension, N-state resonances appear as narrow lines surrounded by negative tails, while I-state resonances appear as broadened lines. The reconstruction of the pure I-state spectrum implies the removal of the N-state contribution to the real time spectrum, taking into account this modification of the line shape of the peaks. This is done considering the kinteic evolution of the native state as an apodization function, as explained below. Line broadening for the I-state resonances can be reduced using experimental hints. First, theory shows that the line broadening induced on ^{13}C resonances is more important than for the ^{15}N resonances recorded in constant-time mode, therefore ^{13}C is sampled using the second dimension (t_2). Second, a constant time scheme can be used to further reduce the line broadening of the ^{15}N resonances. These two experimental methods show sufficient efficiency to proceed with the different analyses shown in sections 13.2 and following.

Method summary: First a 3D spectrum is recorded during the refolding process (the real-time spectrum). The experimental time of this 3D data acquisition has to match the lifetime of the I-state; in other words, the acquisition should stop once the intensity of the I-state signals reaches the noise level. To remove the contribution of the N-state resonances to the real-time spectrum, a second 3D spectrum is acquired under steady-state conditions (steady-state spectrum), i.e. immediately after the end of the folding process. This data set is typically acquired over a much longer time period in order to increase the signal-to-noise ratio (S/N). During data processing, an apodization function, corresponding to the folding kinetics of the N-state, is applied to the steady-state, pure N-state spectrum. Finally, a pure I-state spectrum, devoid of any N-state contributions, is obtained as the difference between the real-time and the apodized steady-state spectra. Kinetics-induced line broadening in the I-state spectra can be limited by choosing short scan times and long maximal evolution times in the last incremented dimension.

13.2 Assigning a transient state

13.2.1 Projected 3D HNCO experiment

The HNCO has been chosen as the first 3D experiment to be tested because it is the most sensitive of the common assignment experiments. Since a standard HNCO requires at least one hour of experimental time, the standard scheme cannot be used to study the intermediate state of B2M. several strategies have been developed over the years to make the acquisition of 3D spectra faster, such as spectral aliasing or data sampling.

As seen in section 7.3, projection NMR allows the acquisition of the information contained in a 3D experiment using a much less time consuming set of 2D experiments, each of which being performed within a few minutes. This allows the repetition of the acquisition along the folding, through a cumulative discrete acquisition scheme, as described in section 13.1.1.

A projection scheme using three projection angles was used to assess for the 3D HNCO experiment: $+\pi/4$, $-\pi/4$, 0. The 0 angle corresponds to a standard 2D-HN experiment. Following the general theory exposed in aforementioned section, on projection planes $+\pi/4$, $-\pi/4$ and 0 the observed indirect chemical shift is given by equation (13.1). This shows clearly the easy analysis based on a central peak (0) and two symmetrical peaks ($+\pi/4$,$-\pi/4$) that contain the CO chemical shift information.

$$\begin{cases} w_{+\pi/4} = w_N + w_{CO} \\ w_{-\pi/4} = w_N - w_{CO} \\ w_0 = w_N \end{cases} \tag{13.1}$$

But as seen on figure 13.3, instead of the detection of most of the CO groups of the protein, the signal to noise ratio allows the detection of only a few peaks. Since the HNCO experiment is the most sensitive of the experiments used to assign a protein, the assignment of B2M's intermediate through projected HNCA or HNCB is out of reach using this cumulative technique.

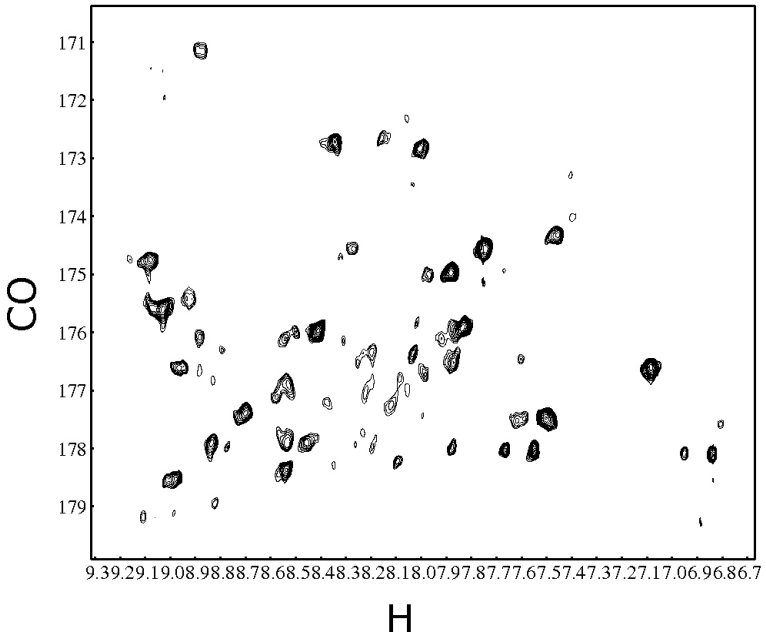
13.2.2 Continuous HNCO and HNCA

The strategy outlined in section 13.1.2 was first applied to HNCO, the most sensitive 3D triple-resonance experiment. At high magnetic field strengths (≥800 MHz), BEST-TROSY generally outperforms BEST-HSQC in terms of sensitivity and spectral resolution[60]. The $^{15}N{-}^{13}CO$ projections of the real-time, N-state, and reconstructed I-state 3D BT-HNCO spectra recorded at 800 MHz ^{1}H frequency are shown in Figure 13.4(b). The quality of these HNCO data was sufficient to detect a similar number of correlation peaks (71) as observed in the I-state $^{1}H{-}^{15}N$ SOFAST-HMQC spectrum (Figure 13.5). The next step was to record a BT-HNCA spectrum required for sequential NMR assignment. This experiment was performed at a magnetic field strength of 1000 MHz in order to further boost the experimental sensitivity. An average SNR increase of $\sim 50\%$ was observed for BT-HNCA when comparing steadystate data recorded at 800 and 1000 MHz (Figure S2).

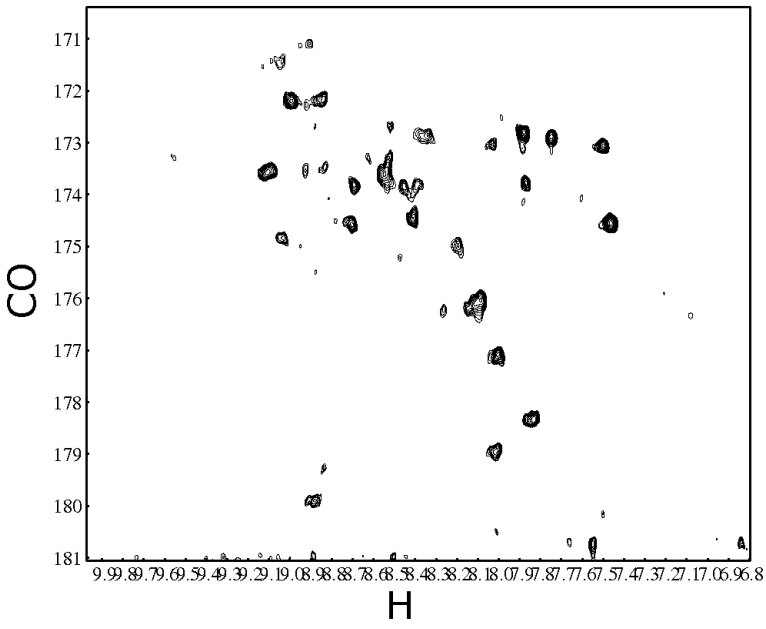
The HNCA data (Figure 13.7) allowed unambiguous sequential assignment of 63 H−N moieties out of the 95 non-proline B2M residues. This corresponds to the majority of correlation peaks detected in the 1H−15N correlation spectrum (Figure 13.5). Interestingly, the peptide regions encompassing Met^{1}-Lys^{6}, Val^{27}-Glu^{36}, Ser^{53}-Lys^{58}, and Phe^{62}-Leu^{65} could not be assigned, because they did not give rise to detectable NMR signals in the 3D spectra, most likely due to extensive line broadening as shown in section 13.3.

13.3 Chemical shift analysis

The chemical shift dispersion observed in the $^{1}H{-}^{15}N$ correlation spectrum of the I-state (Figure 13.5) provides a first indication that the overall structure of this transient state is "native-like". This is further supported by the ^{13}CO and ^{13}CA secondary chemical shift (SCS) data, calculated as the difference between the measured chemical shifts and tabulated random coil values corrected for next neighbor effects[149]. These SCS data are sensitive reporters of local structure (backbone dihedral angles). In particular, a series of consecutive negative SCS values are indicative for β-strand structures. Comparison of measured values for the N- and I-states (Figure 13.8 (a)) shows that the location and population of β-strands in B2M is very similar in these two states. The only remarkable difference is observed for strand D, where the SCS



(a)



(b)

Figure 13.3: Projected HNCO spectra of the I-state with angle $\pi/4$(a) and $-\pi/4$ (b), using the projection BEST-HNCO pulse scheme.

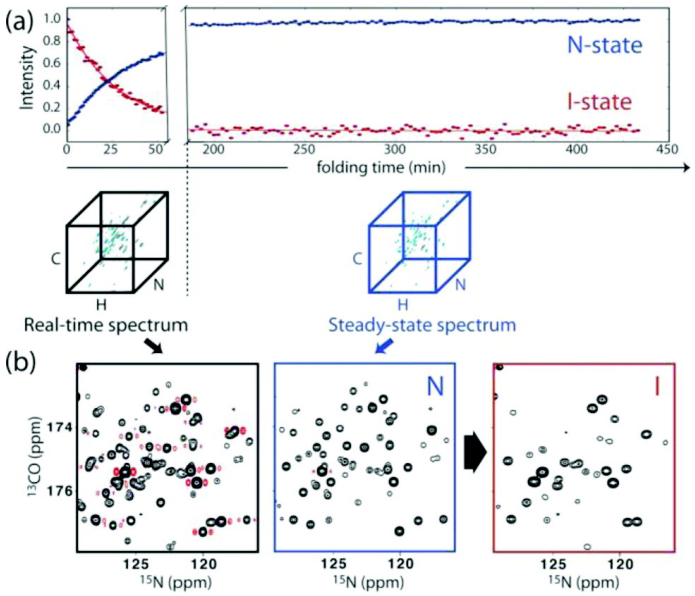


Figure 13.4: Strategy used to reconstruct pure I-state 3D spectra. (a) Kinetic profiles of the N- and I-states are shown during the folding of W60G-B2M, along with the 3D cubes of the acquired HNCO spectra. The red and blue dots are intensities of resolved I- and N-state peaks measured in a series of 2D SOFAST-HMQC spectra recorded during W60G-B2M folding. (b) Projections of the real-time (left), steadys-tate (center), and reconstructed (right) 3D HNCO spectra of B2M. The experimental time for the real-time 3D spectra recorded at 800 MHz for a 0.8 mM 13C,15N-labeled W60G-B2M sample was 40 min.

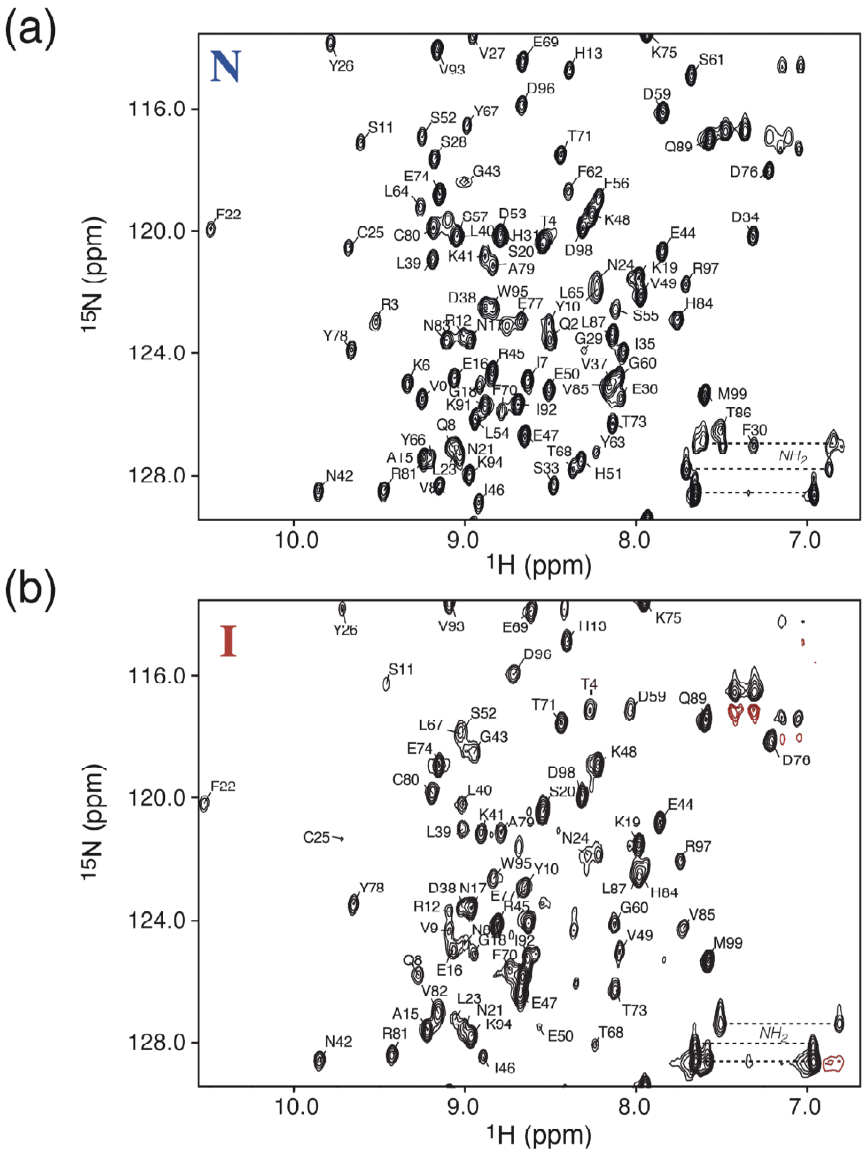


Figure 13.5: Assigned 1H-15N SOFAST-HMQC spectra of (a) N-state and (b) I-state of W60G. Note that some 15N resonances (Tyr26, Val93, Glu69, Asp96 and Val27, Tyr63, Thr68, His51, Thr73, Gly18, Gly29, Phe30, Thr86, Gly43, and Gly60) appear folded in the spectrum due to the reduced 15N spectral width used for this experiment. All assignments were obtained from a real-time 3D BEST-HNCA spectrum, except for Thr4 (I-state) that has been assigned in analogy to previous NMR studies using trifluoroethanol (TFE) that enhances the equilibrium population of the I-state [135].

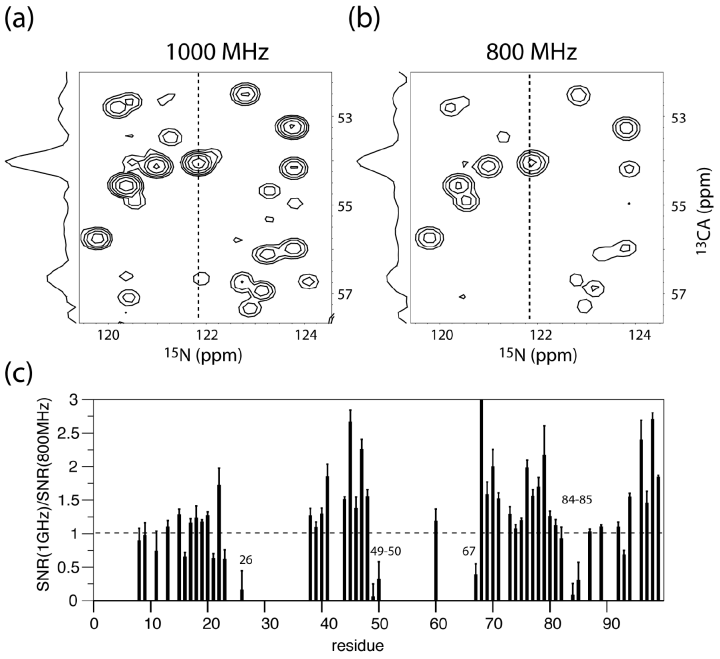


Figure 13.6: 15N-13C projections of 3D BEST-TROSY-HNCA (steady-state) spectra recorded at ^{1}H frequencies of (a) 1000 MHz and (b) 800 MHz. The spectrometer operating at 800 MHz is an Agilent instrument installed in our laboratory (IBS) in Grenoble, while the 1000 MHz instrument is a Bruker spectrometer located at the "Centre de RMN à Très Hauts Champs" in Lyon (France). The vertical dashed lines indicate the position of the 1D slices shown at the left side of each 2D projection. Note that the contour levels are scaled to the corresponding noise levels. By comparison of the intensities of several (N-state) peaks in the two 3D spectra, an average SNR enhancement of 50% was estimated at the higher magnetic field strength. (c) Normalized intensity ratios of I-state signals measured in the reconstructed HNCA spectra recorded at 800 and 1000 MHz.

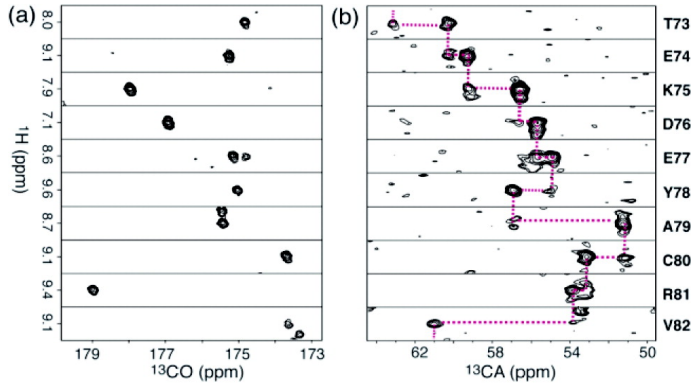


Figure 13.7: Example strip plots extracted from the reconstructed I-state spectra of (a) 3D BT-HNCO and (b) 3D BT-HNCA recorded on a 0.8 mM ^{13}C,^{15}N-labeled W60G sample at 800 MHz (HNCO) and 1 GHz (HNCA). The assignment walk for residues Thr^{73}-Val^{82} is highlighted by dotted (red) lines.

data indicate a less pronounced population of β-strand conformation. In addition to local secondary structural changes, ^{1}H and ^{15}N chemical shifts are also sensitive to hydrogen bonding and tertiary contacts, e.g., to aromatic side chains that induce ring-current shifts. Significant changes ($\Delta\delta$(H,N) $>$ 1.5 ppm) in ^{1}H, ^{15}N chemical shifts between the N- and I-states are found for residues Gln^{8}, Val^{9}, Ser^{11}, Leu^{39}, Val^{49}, Glu^{50}, Asp^{59}, Ser^{61}, Asn^{83}-Val^{85}, and Leu^{87} (Figure 13.8(b)), i.e. generally neighboring unassigned (and undetected) regions of the protein in the Istate. A plot of the observed chemical shift deviations and the NMR-invisible protein regions on the 3D cartoon structure of B2M (Figure 13.8(c)) points out that structural differences between the I- and N-states exist in the apical region close to Pro32, with a particularly strong effect in the second half of the β-sandwich (A, B, D, E), while strands F and G are less affected.

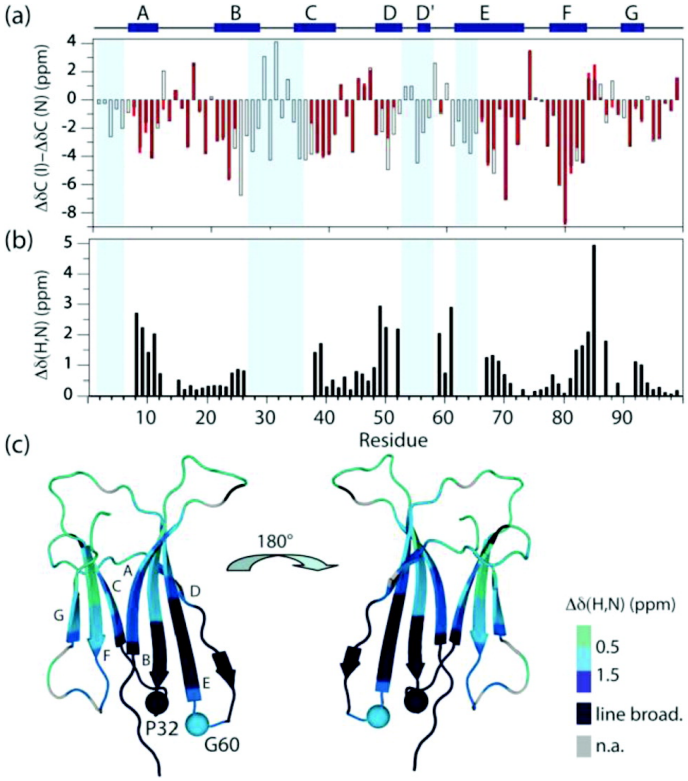


Figure 13.8: (a) Secondary 13C chemical shifts ($\Delta\delta$CO + $\Delta\delta$CA) measured for the N-state (open bars) and I-state (filled red bars) of W60G-B2M. The 1H, 15N chemical shift difference $\Delta\delta HN = [(10\Delta\delta H)^2 + \Delta\delta N^2]^{0.5}$ between I and N is plotted in (b) as a function of sequence and in (c) on the cartoon structure of B2M. NMR-invisible I-state regions are indicated by cyan bars in (a) and (b), and dark blue bars in (c)

13.4 Relaxation measurements in the I state

13.4.1 R_2-BEST-TROSY for relaxation measurements

R_2-BEST-TROSY is a NMR experiment that allows detection of R_2 relaxation rates for all NH groups in the protein. BEST stands for band selective excitation short transient and is based on the use of selective ^1H pulses to enhance longitudinal relaxation, much like SOFAST, see 7.2.3. TROSY stands for transverse relaxation optimized spectroscopy, and is based on the principle of cross correlated relaxation. In-depth theory developments show that two relaxation mechanisms such as dipole-dipole interaction and chemical shift anisotropy not only have specific effects on a spin system, but also lead to positive or negative interferences. This will for example cause the two lines of a scalar-coupled IS spin system doublet to have different line widths due to their different relaxation behavior. The TROSY scheme will select the coherence with the longest relaxation time, therefore yielding the sharpest linewidth. As seen in section 7.1.4, R_2 measurements necessitate the introduction of a spin echo with a variable delay in the experiment. This is made in the R_2-BEST-TROSY pulse sequence using the parameter Δ, as shown on figure 13.9.

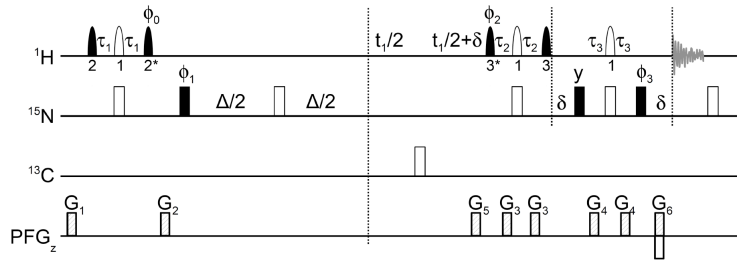


Figure 13.9: R2-BEST-TROSY pulse sequence used for the measurement of R(Nx+2NxHz) relaxation rates (adapted from A. Favier & B. Brutscher, J. Biomol. NMR. 2011, 49, 9-15). Filled and open symbols correspond to 90° and 180° rf pulses, respectively. Amide 1H pulses typically cover a bandwidth of 4 ppm (centered at 8.5 ppm), and have the following shapes and durations at 800 MHz: (1) REBURP, 1.5 ms (δ1), (2) PC9, 2.2 ms (δ2), and (3) EBURP-2, 1.4 ms (δ3). An asterisk indicates time and phase reversal of the corresponding pulse shape. The transfer delays are adjusted to τ1=1/(4 JNH) - 0.5 δ1 - 0.5 δ2, τ 2=1/(4 JNH) - 0.5 δ1 - ϰ δ3 and τ 3=1/(4 JNH) - 0.5 δ1. These settings account for spin evolution during the various shaped 1H pulses. The parameter $\varkappa \approx 0.7$ can be fine tuned to equilibrate the transfer amplitudes of the different coherence transfer pathways for optimal suppression of the unwanted quadruplet components in the spectrum. Pulses are applied along the x-axis unless indicated. The phase φ0 needs to be set to +y or −y, depending on the spectrometer, to add the signals originating from 1H and 15N polarization. Pulsed field gradients are applied along the z-axis (PFGz) with durations (and field strengths) of 300µs (35 G/cm) for G1, 50µs (15 G/cm) for G2, 100µs (30 G/cm) for G3, 1000µs (6 G/cm) for G4, 1000µs (35 G/cm) for G5, 100µs (35 G/cm) for G6. The relative durations of G5 and G6 are fine tuned to fulfill the relation G5/G6 = γH/γN. The phase cycling is: φ1 = −x,x; φ 2 = −y; φ 3 = −x; φ acq = −x,x. For quadrature detection in T_1, echo-antiecho data are recorded by inverting the sign of gradient G6 together with phases φ 2 and φ 3.

13.4.2 HET-SOFAST

HET-SOFAST[146] uses longitudinal relaxation properties of proton spins for the fast detection of structural compactness and heterogeneity along the polypeptide backbone of proteins.

An additional band selective inversion pulse is added to the standard SOFAST-HMQC pulse sequence at the beginning of the recycle period or contact time t_c on the H_{sat} proton channel (figure 13.10). The experiment measures the effect of this perturbation (saturation) on the spin polarization of the amide protons after a contact time t_c. Two data sets are recorded, with and without the H_{sat} inversion pulse. The ratio of the peak intensities ($= I_{sat}/I_{ref}$) measured in the two spectra then provides insight into the interaction between amide protons (H_N) and saturated protons (H_{sat}).

Two types of proton interactions may be monitored by the HET-SOFAST experiment. If the H_{sat} pulses are applied to the aliphatic protons, the experiment measures the $^1H-^1H$ spin diffusion or NOE effect (λ_{noe}). If the water polarization is inverted, the

intensity ratios (λ_{noe}) provide a measure of the amide water hydrogen exchange rates. The two observables, λ_{noe} and λ_{ex} yield complementary information on the structure and mobility of the polypeptide chain. The λ_{noe} values report on the local ^1H spin network, i.e. the number, density, and mobility of protons close to a particular amide, whereas the λ_{ex} values are indicative of the water accessibility of the amide hydrogens.

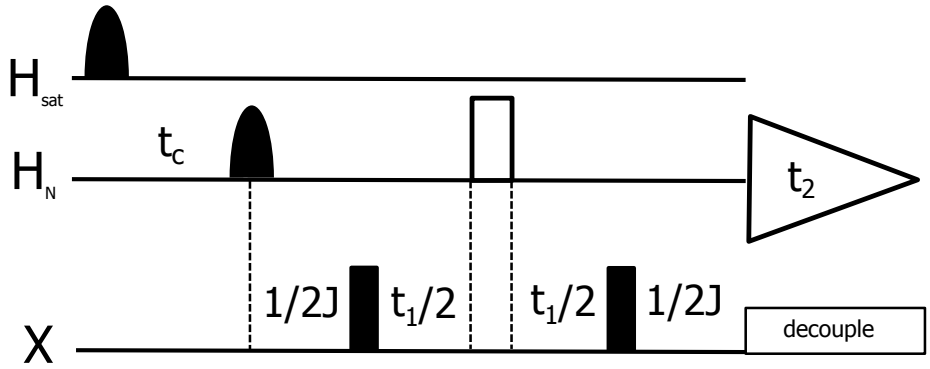


Figure 13.10: HET-SOFAST pulse sequence, similar to the SOFAST pulse sequence with an added selective saturation pulse at the beginning

13.4.3 Results

The missing NMR signals for several consecutive regions that are structurally close to Pro32 indicate the presence of conformational dynamics in this part of the protein. Toward these regions, the NMR signal decreases gradually, as nicely seen from the plot of intensity ratios of I- versus N-state peaks in HNCO spectra (Figure 13.11(a)). To obtain additional, more quantitative information on the local conformational dynamics in the I-state of B2M, we performed two simple spin relaxation measurements during protein refolding. Such measurements require the recording of a small number of 2D 1H−15N correlation spectra (pseudo-3D experiment) using different parameter sets, e.g., varying a relaxation delay. The total acquisition time for such a pseudo-3D experiment is generally only a few minutes, much shorter than the protein folding time. Therefore, data acquisition is repeated several times during realtime folding.

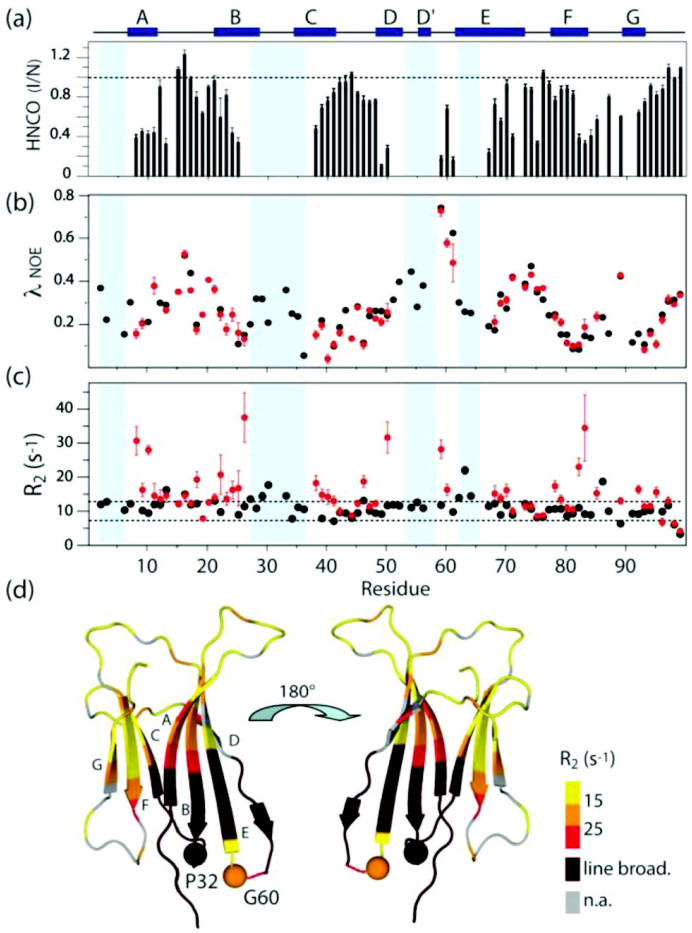


Figure 13.11: (a) Ratios of peak intensities measured in I-state and N-state 3D HNCO spectra. (b) λNOE values obtained from HET-SOFAST experiment. (c) R2(Nx + 2NxHz) relaxation rates measured with R2- BEST-TROSY sequence (Figure S4). N- and I-state values are plotted in black and red, respectively. Both experiments were performed at 800 MHz on a 0.6 mM 13C,15N-labeled W60G-B2M sample. In (d) the measured R2 values are color-coded on the cartoon structure of B2M. NMR-invisible I-state regions are indicated by cyan bars in (a)– (c), and dark red in (d).

First, a HET-SOFAST experiment[146] was performed that exploits the difference in amide ^1H T_1 relaxation in the presence/absence of aliphatic ^1H saturation to probe the structural compactness (local ^1H density) and sub-ns time scale dynamics. The intensity ratios (λ_{NOE} = Isat/Iref) measured for individual amide protons in the N- and I-states of W60G are compared in the graph of Figure 13.11(b). Low λ_{NOE} values are indicative of rigid, compact local structure, while λ_{NOE} values close to 1 are expected for highly flexible loop regions. From these data we can conclude that, within the experimental error, the two states behave very similarly in terms of local structural compactness and fast time scale dynamics, even in the regions with significant differences in ^1H , ^{15}N chemical shifts (Figure 13.8). Of course, no

conclusion can be drawn for the protein regions that are NMR invisible.

In a second experiment, we measured the transverse relaxation rates of the ^{15}N coherence $N_x + 2N_xH_z$ (TROSY line) by inserting a variable spin−echo delay into the ^1H $-^{15}N$ BEST-TROSY experiment (see section 13.4.1). This R2-BEST-TROSY experiment yields information on both fast motions (ps−ns), comprising the molecular tumbling and local bondvector fluctuations, and conformational exchange dynamics occurring on the µs−ms time scale. Relaxation rates were obtained by fitting the cross-peak intensities measured for three relaxation times T (1, 20, and 60 ms) to the function $I(T) = A\exp(-R_2T)$ (Figure 13.12). The measured rates are shown in Figure 13.11(c). A first interesting conclusion from these data is that, for protein regions that do not show significant chemical shift variations between I- and N-states, the measured relaxation rates are also very similar ($\sim 10s^{-1}$). This observation indicates similar molecular tumbling correlation times for the I- and N-states, excluding the hypothesis of a fast monomer−oligomer exchange process as a possible explanation for the missing I-state correlation peaks. This is consistent with a monomer-oligomer equilibrium occurring on timescales of 1 to 10 s^{-1}, as described in section 10.3.2.

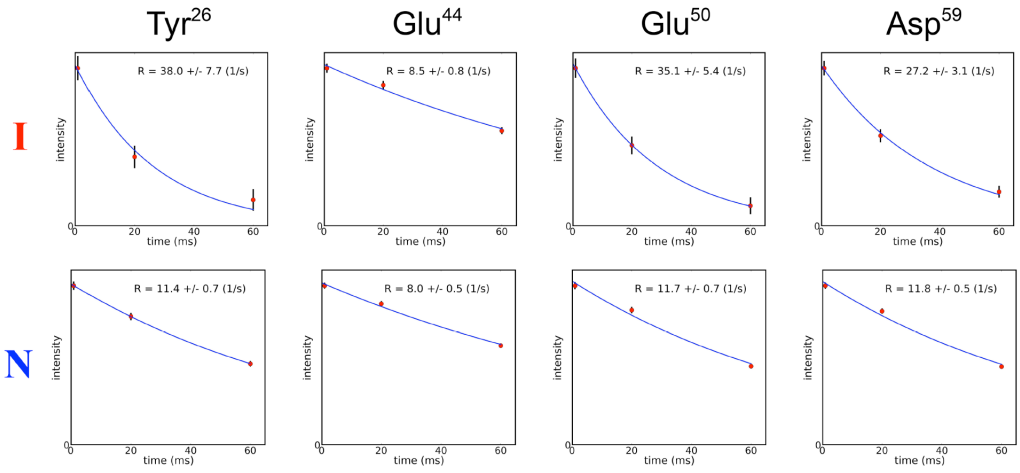


Figure 13.12: Peak intensity decays measured for selected residues in the Istate (top) and N state (bottom) of W60G-B2M. Three residues (Tyr26, Glu50, Asp59) show faster transverse relaxation in the I-state than in the N-state, while the relaxation rates measured for a fourth residue (Glu44) are approximately equal in both states.

Most interestingly, transverse relaxation in the I-state is increased in regions close to the NMR-invisible residues, further supporting the idea that the absence of NMR signal is due to conformational exchange processes on the µs−ms timescale in the apical side of the protein close to Pro32 (Figure 13.11(d)). Based on the observed profiles of chemical shift changes and 15N transverse relaxation rates, it is tempting to speculate that this conformational exchange process involves a collective opening motion in the apical half of B2M that is destabilized by the non-native transprolyl

bond.

13.5 Conclusions

In summary, the results of this chapter demonstrate the ability of sensitivity-enhanced, fast real-time 3D NMR spectroscopy combined with high-field NMR instruments to provide insight into the structural and dynamic properties of a protein folding intermediate of B2M at atomic resolution. In particular, we have shown that, once sequential NMR assignment has been obtained, quantitative information on local dynamics at various time scales can be obtained from real-time spin relaxation measurements. This approach can be readily applied to other biologically relevant systems and provides a powerful strategy for NMR-based characterization of protein states with half-lifetime of a few tens of minutes. As demonstrated here, fast real-time 3D NMR represents a unique method for the characterization of "long-lived" transiently populated protein states. It may also prove useful for in-cell NMR studies[150] where the protein of interest has often a similar short lifetime, and where structural changes occurring after some cellular stimulus may be monitored by real-time NMR methods.

If the speculated opening motion conformational exchange in the apical half of W60G is attractive, the question remains of the extension of this result to the WT protein. As seen in section 8.3, the folding mechanism of W60G is different from that of WT B2M. Especially, it populates the oligomeric states of B2M folding with a smaller fraction. Yet, does the monomeric state I_1 of WT B2M also show this conformational exchange? Is it linked to the oligomerization equilibrium that occurs during folding? Only the study of WT B2M itself would allow such a confirmation. This study will push even further the need of optimization of the NMR methods presented in this chapter, as with a lower population of I_1 goes an increased need of sensitivity.

Chapter 14

Conclusion and Perspectives

B2M is both a fruitful and difficult object of study. B2M production is complex, requiring optimization to obtain a well folded protein and reach yields suitable to NMR and SAXS studies. B2M folding is highly sensitive to buffer, temperature, concentration and often preparation conditions. Yet our studies, using several biophysical methods, revealed essential features of the folding mechanism and of the structure and dynamics of folding intermediates of this protein.

A first outcome of our studies is that folding and oligomerization are co-existing processes. A major finding is the existence of a monomer-oligomer equilibrium between I_1 and I_2 folding intermediate states. Indirectly detected using real time NMR methods like SOFAST, I_2 was directly detected and characterized using SAXS: I_2 is likely to be a dimer. Folding intermediate states of B2M had been shown to favor fibril formation: this is easily explained by the existence of a dimeric folding intermediate state with a significant population.

A combination of biophysical methods allowed the characterization of this monomer-oligomer equilibrium. Using SAXS, and later confirmed by NMR relaxation experiments, stoichiometry is shown to be a monomer-dimer equilibrium. Hints on the timescale of this equilibrium can also be defined using real time NMR experiments: the equilibrium is fast compared to the dead time of the experiment, even though the available data can not discriminate between a fast or a slow exchange relatively to the NMR timescale. Further work based on the methodology applied to the folding of the W60G-B2M mutant, including a further optimization of the sensitivity of the experiment, will give a sharper picture of the I_1-I_2 equilibrium of the WT protein, and may provide information on the timescale of the equilibrium.

The thorough study of the folding of B2M pushes biophysical methods to their limits: sensitivity and acquisition time for NMR, polydispersity for SAXS. Yet in both cases a large oligomer (I_3) that disappears within minutes was detected, and confirmed

using UV-fluorescence. Characterization of I_3 will demand further methodological developments, a new experimentation plan including a full dilution scale, or double jump experiments for example. Another possible strategy is the use of a perdeuterated methyl protonated samples of B2M for NMR experiments, that may lead to a direct characterization of I_3. A question that arises is the comparison of this large oligomer and oligomeric intermediate states that are populated during the formation of fibrils. Other biophysical methods, such as ESI mass spectroscopy, may be an interesting input.

Tackling the limits of biophysical methods leads to methodological developments. For example, to study the structure and dynamics of I_1, the continuous data acquisition method allowed the assignment of this species that has a half-lifetime of ~ 20 minutes. A conformational exchange was discovered for the I_1 state of the W60G-B2M mutant, through the development of a spin relaxation measurement experiment: R_2-BEST-TROSY. The methods developed for this study may be later used to study the folding and folding intermediate states of other proteins, or in other contexts in which the short lifetime of the protein is an issue, as for in-cell NMR experiments.

Our studies are of course far from an application or a concrete result in the fight against misfolding diseases such as dialysis related amyloidosis or Parkinson's. But the discovery of oligomeric folding intermediate states underlines that oligomerization (including fibril formation) and folding should not be studied separately, and are processes that are closely related. Methodological developments included in our work can be applied to other proteins as well as other contexts. Hopefully these questionings and developments will constitute a step forward toward a better understanding of such diseases.

Part VI

Annex

Publications based on the studies described in this manuscript

- **Thomas Cutuil***, Enrico Rennella*, Isabel Ayala, Paul Schanda, Frank Gabel, Vincent Forge, Alessandra Corazza, Gennaro Esposito, Bernhard Brutscher. Oligomeric species accumulate during the early stages of β_2-microglobulin refolding. Manuscript in preparation.

- Enrico Rennella, **Thomas Cutuil**, Paul Schanda, Isabel Ayala, Vincent Forge, and Bernhard Brutscher. Real-time nmr characterization of structure and dynamics in a transiently populated protein folding intermediate. J Am Chem Soc, 134(19):80668069, May 2012.

- Alessandra Corazza, Enrico Rennella, Paul Schanda, Maria Chiara Mimmi, **Thomas Cutuil**, Sara Raimondi, Soa Giorgetti, Federico Fogolari, Paolo Viglino, Lucio Frydman, Maayan Gal, Vittorio Bellotti, Bernhard Brutscher, and Gennaro Esposito. Native-unlike long-lived intermediates along the folding pathway of the amyloidogenic protein beta2-microglobulin revealed by real-time two-dimensional nmr. J Biol Chem, 285(8):58275835, Feb 2010.

Résumé de la thèse

Remarque: ce résumé est présent à titre d'information, et n'est pas à considérer comme une source.

Le repliement des protéines est le procédé par lequel une protéine atteint une conformation stable et fonctionnelle. Le principe d'Anfinsen dit que cette conformation est unique, et déterminée uniquement par la séquence d'acides-aminés de la protéine, faisant ainsi le lien entre l'information génétique codée dans l'ADN et la fonction agissante de la protéine via une interaction clé-serrure et donc une structure tridimensionnelle spécifique. Toutefois cela doit aujourd'hui être nuancé, la conformation des protéines devant être considérée plutôt comme un ensemble d'états similaires, plutôt que comme une structure rigide et statique. De plus, des protéines intrinsèquement désordonnées sont aujourd'hui découvertes, montrant que la fonction peut ne pas être liée à une structure bien définie. L'acquisition d'une structure tertiaire pourra même être obtenue lors de la liaison avec une autre protéine ou un substrat.

La vision traditionnelle du repliement des protéines dérive du paradoxe de Levinthal, qui dit que si une chaine polypeptidique devait de manière séquentielle expérimenter toutes les conformations possibles avant d'atteindre, par hasard, sa conformation fonctionelle, cela impliquerait un temps de repliement plus long que l'age de l'Univers. Et pourtant, la plupart des petites protéines se replient en quelques millisecondes. Ce paradoxe suggère que les protéines suivent un chemin préférentiel de repliement. Cela peut être illustré par un paysage énergétique de repliement en entonnoir, qui peut contenir des minima locaux correspondant à des états intermédiaires de repliement.

La principale limite à l'étude du repliement des protéines est son échelle de temps : la microseconde ou la milliseconde pour la plupart des petites protéines. Les méthodes traditionnelles d'étude des structures des protéines ont des résolutions temporelles de l'ordre de la seconde dans le meilleur des cas. Les solutions traditionnelles à ces difficultés sont la baisse de la température, qui allonge le temps de repliement, et les développements méthodologiques.

Les protéines mal repliées sont habituellement détectées par la cellule et dégradées par protéolyse, mais certaines maladies sont dues à des protéines ayant subi une mutation, qui ne peuvent atteindre leur conformation fonctionnelle, ou à des protéines qui ne peuvent atteindre leur conformation fonctionnelle dans certaines conditions. Par ailleurs, un dysfonctionnement du système protéolytique peut aussi entrainer des pathologies.

Les amyloses sont un groupe défini de conditions médicales dans lesquelles des protéines normalement solubles et innoffensives polymérisent pour former des fibrilles insolubles, et apparaissent donc dans la catégories des maladies du mauvais repliement. Les fibrilles amyloides grandissent et s'associent aux constituants extracellulaires,

comme le plasme ou des protéoglycanes, pour former des dépôts amyloides dans les tissus. Les maladies d'Alzheimer, Parkinson, certains cancers, ainsi que l'amylose liée à la dialyse présentent des dépots amyloides. Cette dernière est en particulier liée à la polymérisation de la Beta-2-microglobuline, protéine du système immunitaire, qui peut dans des conditions physiologiques être soluble et circulante.

De nombreuses techniques permettent l'étude du repliement des protéines, sous un angle soit cinétique, soit structurelle. L'étude du repliement suppose l'étude d'échantillons complexes, incluant l'existence d'états transitoires, une polydispersité certaine, une propension à l'agrégation, différentes échelles de taille et de temps, ainsi que des cinétiques rapides. Une étude sérieuse du repliement des protéines suppose donc l'utilisation de méthodes complémentaires. On citera la spectroscopie UV-visible, le dichroïsme circulaire, la spectroscopie à infra-rouge, la diffraction X aux petites angles (SAXS), et la RMN. La b2m est une protéine utilisée comme modèle depuis de nombreuses années, en raison de trois caractéristiques. D'abord la B2M forme des fibrilles amyloïdes à la fois in vivo et in vitro. Ensuite ces fibrilles amyloïdes sont impliquées dans une maladie appelée amylose liée à la dialyse, leur donnant une importance médicale directe. Enfin alors que les protéines de petite taille tendent à se replier en quelques millisecondes, la B2M se replie en une douzaine de minutes dans des conditions physiologiques, ce qui permet d'étudier son repliement avec des méthodes biophysiques usuelles.

La formation des fibrilles amyloides de B2M est étudiée depuis 1985, principalement via diverses déstabilisations de la protéine, comme des mutations ou troncatures, interaction avec des ions cuivre, des solvants organiques, du collagène, ou des contraintes physico-chimiques comme la température, une concentration élevée de sel, ou une agitation importante. Tous ces changements ont une influence sur l'équilibre cis/trans du lien peptidique entre His31 et Pro32. Ce lien est dans une conformation cis dans l'état natif, alors qu'une conformation trans est observée dans les fibrilles. D'autres régions de la protéine, comme la boucle DE, ont été désignées comme participant à la polymérisation de la B2M.

La formation des fibrilles amyloides passe pour B2M comme pour la plupart des protéines par la formation d'oligomères. L'utilisation de DTT a permis de mettre en évidence la formation d'une échelle d'oligomères de différentes tailles, menant à la formation de fibrilles amyloides. Par ailleurs, l'utilisation d'anticorps à domaine unique a aussi permis d'obtenir une structure crystallographique d'un dimère de la B2M, annoncé comme un intermédiaire sur le chemin de la formation des fibrilles amyloides. Enfin, l'utilisation d'ions cuivre a permis d'établir un changement conformationnel de B2M impliqué dans une oligomérisation puis la formation de fibrilles amyloides.

Le repliement de la B2M a principalement été étudié dans la littérature par dichroïsme circulaire et par spectroscopie UV-visible. Les premières expériences par Chiti et al en 2001 ont montré que le repliement de la B2M présentait trois étapes. Tout d'abord

l'état déplié était très rapidement transformé en un premier état intermédiaire, avec un temps caractéristique inférieur au temps mort de l'expérience, à savoir quelques millisecondes. Puis ce premier état intermédiaire donne naissance à un deuxième état intermédiaire, en moins de trois secondes. Puis une troisième phase de repliement, lente, est terminée en une dizaine de minutes, pour atteindre l'état natif. Ce modèle de repliement est linéaire. La phase rapide de trois secondes n'est pas détectée par dichroisme circulaire, ce qui suggère que les deux états intermédiaires ont des structures similaires pour ce qui concerne cette méthode.

L'utilisation des mêmes méthodes par Jahn et al en 2006, en ajoutant une série de conditions de repliement (faisant varier la concentration finale de Gd-HCl), fait apparaître un schéma de repliement parrallèle comprenant cinq états. Il n'est toutefois pas incompatible avec le modèle de Chiti, en raison de la faible population des états aux moments précis qui pourraient discriminer entre les deux modèles. Une troisième étude par Sakata et al en 2008 conclut, par l'utilisation de méthodes similaires et des expériences de double saut, à un schéma à quatre états. Le repliement de la B2M est donc complexe, et suppose une analyse fine des données recueillies.

Le fait majeur du repliement de la B2M est la transition trans vers cis du lien peptidique His31-Pro32. Cette transition étant impliquée dans l'oligomérisation de la B2M et dans la formation de fibrilles, il apparaît possible que les deux phénomènes soient liés.

Les développements récents de la résonnance magnétique nucléaire (RMN) ont permis de réduire le temps nécessaire à l'acquisition de données. Plusieurs méthodes peuvent être mises à contribution, comme l'acquisition parcellaire, ou l'utilisation des propriétés de relaxation notamment longitudinale. Cela a mené à la conception des méthodes BEST et SOFAST, méthodes permettant d'acquérir des spectres 2D en quelques secondes par exemple. La RMN permet d'obtenir de nombreuses informations sur la structure, la dynamique, des protéines, mais aussi sur les phénomènes cinétiques. Elle permet notamment d'évaluer l'échelle de temps de phénomènes dynamiques ou cinétiques, mais aussi de détecter les parties de la protéine impliquées dans tel ou tel changement structurel.

La méthode de base de ces études est le repliement utilisant un saut de pH à partir d'un état acide dénaturé de la B2M. Ce saut de pH est provoqué par un injecteur permettant de réduire le temps mort des expériences RMN à quelques secondes. La B2M utilisée a été obtenu par production recombinante, et marquage isotopique simple (N15) ou double (C13-N15). L'évolution cinétique du repliement est suivie par des acquisitions successives, environ toutes les quinze secondes.

Le suivi du repliement de la B2M fait apparaître plusieurs informations. Tout d'abord, juste après le saut de pH, le spectre acquis ne montre aucune trace de l'état acide dénaturé. Cet état a laissé place à des états dont le spectre est typique d'espèces au

moins partiellement repliées, au cours du temps mort de l'expérience. Cela correspond à la phase immédiate (burst phase) et à la phase rapide évoquée dans la littérature. Il est possible sur ces spectres d'isoler trois classes de pics : ceux correspondant à l'état natif, ceux montrant une superposition de deux états, et ceux typiques d'une espèce intermédiaire. L'intensité mesurée pour les pics de l'état natif est de 10% de la valeur à la fin du repliement. On retrouve ensuite la phase lente précédemment décrite, avec passage de l'état intermédiaire It à l'état natif N. Toutefois, la classe des pics comportant une superposition I et N ne donne pas lieu à une intensité constante au cours du repliement. Cela signifie qu'un autre état intermédiaire au moins, invisible en RMN, est peuplé de manière non négligeable au cours du repliement.

Cela permet de construire de nouveaux modèles de repliement. Toutefois la question de la caractérisation de l'état intermédiaire invisible en RMN reste posée. Plusieurs hypothèses peuvent expliquer cette invisibilité. Soit cet état est oligomérique, donc de taille trop importante pour être détectable, soit cet état est trop dynamique pour être détecté. Les modèles obtenus par RMN peuvent être considérés comme compatibles avec les modèles de la littérature.

Cet état intermédiaire invisible peut cependant être étudié par RMN. En effet ce manque initial d'intensité est un témoin de la population de cet état. En faisant varier les conditions expérimentales, on peut voir que la population de l'état intermédiaire invisible augmente avec la hausse de la concentration et la baisse de la température. Cela laisse supposer que cet état intermédiaire invisible en RMN, et qui n'avait pas été décrit dans la littérature, serait un oligomère.

Cet état intermédiaire invisible par RMN est en équilibre avec l'état intermédiaire visible. Plusieurs hypothèses apparaissent quant à cet équilibre. Soit l'échange est lent comparé à l'échelle de la RMN, et alors l'état invisible est soit un grand oligomère, soit un petit oligomère dont le spectre est exactement superposé à l'état intermédiaire monomérique. Soit l'échange est rapide, et l'état invisible est un petit oligomère.

L'utilisation d'autres méthodes biophysiques permet d'aller plus loin dans la caractérisation de cet état invisible par RMN. La plus importante d'entre elles, puisqu'il s'agit de détecter et de caractériser des oligomères, est la diffraction de rayons X aux petits angles (SAXS). Cette technique peut être utilisée en temps réel, pour suivre le repliement de la B2M, avec des contraintes méthodologiques liées à la consommation de protéine, réduisant le nombre de données mesurées au cours du repliement à moins d'une dizaine. Le temps mort est par ailleurs plus long qu'en RMN, de l'ordre d'une à deux minutes.

Le suivi du repliement par SAXS montre que le poids moléculaire moyen dans l'échantillon diminue au cours du repliement. Cela montre directement la présence d'oligomère au cours du repliement, dont la population diminue pour aller vers l'état natif. La cinétique globale observée est similaire à celle obtenue en RMN. Le traitement des données

SAXS permet par ailleurs de dégager deux types d'oligomères : l'un disparaissant en quelques minutes, l'autre suivant la cinétique observée en RMN. Le premier oligomère est supposé être un grand oligomère, alors que le second est très probablement un dimère, selon les données SAXS.

Une autre méthode biophysique, à savoir la spectrofluorimétrie UV en présence de thioflavine T, permet de caractériser encore plus précisément ces états intermédiaires oligomériques. Ainsi la thioflavine T, qui se lie aux structures similaires à des fibrilles amyloides, se lie au premier oligomère détecté en SAXS et qui disparaît en quelques minutes, alors qu'elle ne se lie pas à l'état intermédiaire dimérique. Cela montre que le grand oligomère possède probablement une structure proche de celle rencontrée dans les fibrilles amyloides.

En cherchant spécifiquement ce grand oligomère, avec l'échelle de temps concernée, il est en fait possible de détecter trois pics spécifiques dans les spectres RMN obtenus lors du repliement. Ces pics correspondent probablement à quelques acides aminés flexibles, probablement à l'une des terminaisons de la protéine.

Ces découvertes posent toutefois la question de la détection de l'état intermédiaire dimérique par RMN. En effet un dimère d'une protéine comme la B2M devrait être détectable directement par RMN. Cette caractérisation par SAXS permet ainsi d'éliminer l'hypothèse d'un échange lent entre état intermédiaire oligomérique de grande taille et état intermédiaire monomérique. Toutefois les données ne sont pas suffisantes pour trancher entre l'hypothèse d'un échange rapide ou d'un échange lent avec superposition exacte.

Les données RMN et SAXS peuvent être modélisées simultanément pour aboutir à un modèle de repliement à trois états intermédiaires : un état monomérique, un état dimérique ou oligomérique de petite taille, et un état oligomérique de grande taille à structure fibrilloide.

Les états d'équilibre (natifs) de la B2M et de mutants de la B2M peuvent être étudiés par SAXS. Cette étude montre que le mutant P32A peut constituer un modèle d'équilibre des états intermédiaire, présentant une oligomérisation importante dans des conditions physiologiques. Par ailleurs, le mutant W60G apparaît comme un modèle de l'état natif monomérique de la B2M. Les données SAXS montrent en effet que même dans des conditions physiologiques, la B2M peuple de manière non négligeable des états oligomériques.

D'autres méthodes RMN permettant la caractérisation de l'état intermédiaire monomérique ont pu être développées, permettant l'attribution de cet état au cours du repliement de la B2M, grâce à une forte réduction du temps d'acquisition nécessaire, et la mise en place d'une acquisition dite continue. L'attribution permet d'attribuer au résidu correspondant les évolutions de déplacement chimique au cours du repliement. Le développement de méthodes rapides d'étude de relaxation RMN permet par ailleurs

de montrer qu'il existe un certain échange conformationnel pourrait expliquer le fait que certaines parties de la protéine dans l'état intermédiaire monomérique restent invisibles par RMN. Le développement de toutes ces méthodes montre que la RMN permet d'obtenir des données sur la structure et la dynamique d'états transitoires d'une protéine.

La B2M est un objet d'étude complexe mais riche en enseignements. L'utilisation de techniques biophysiques complémentaires montre tout son intérêt dans ce cadre. La découverte de multiples intermédiaires de repliement et leur caractérisation est l'avancée majeure de ce travail, montrant les liens entre oligomérisation et repliement. Par ailleurs, les méthodes développées ici pourront être appliquées à d'autres protéines.

NMR: folding of B2M detailed data

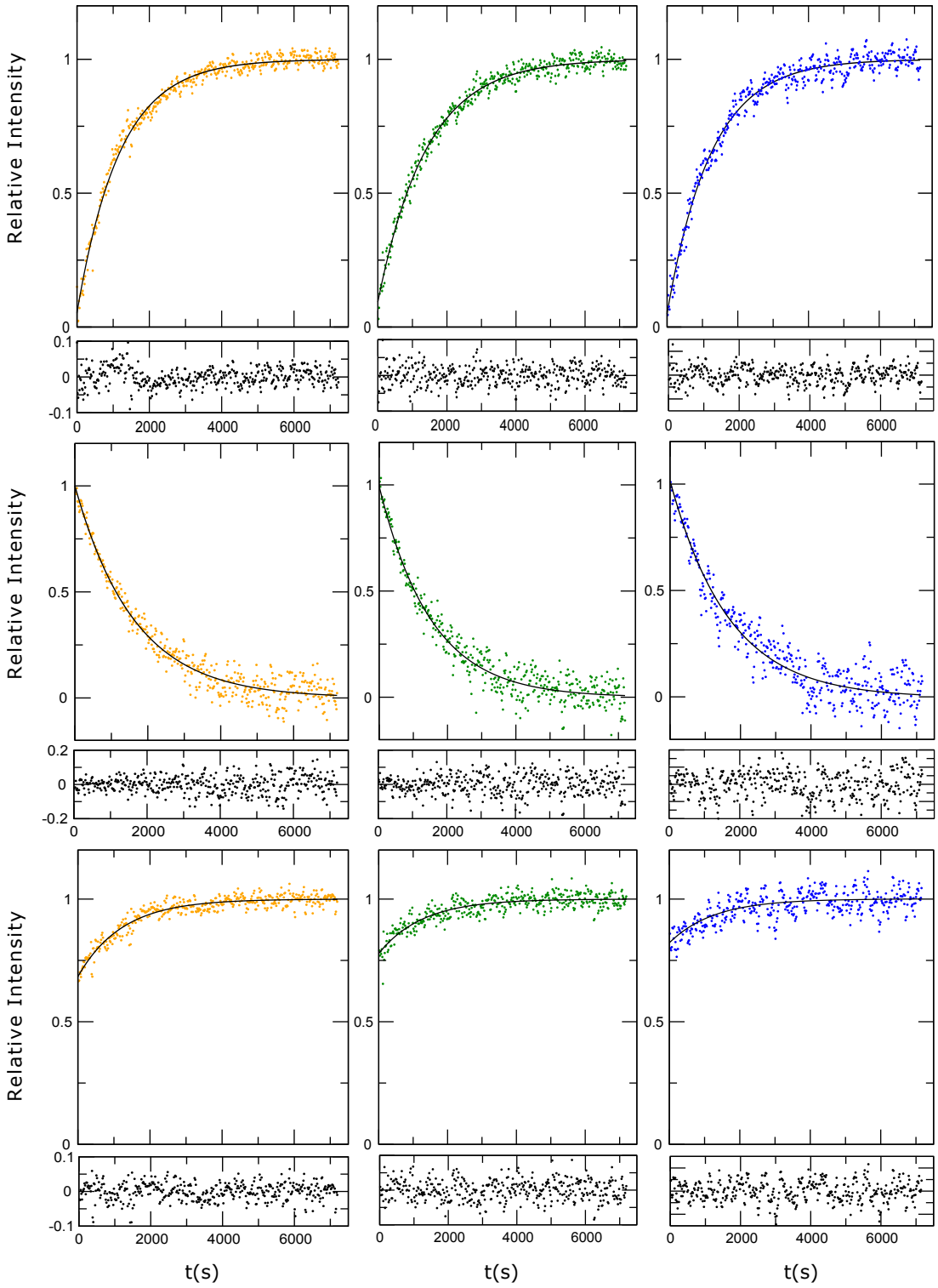


Figure 14.2: Folding of WT-B2M at 20°C. Top line: N class. Center line: I class. Bottom line: I+N class. Left column: 750µM B2M. Center column: 375µM B2M. Right column: 188µM B2M.

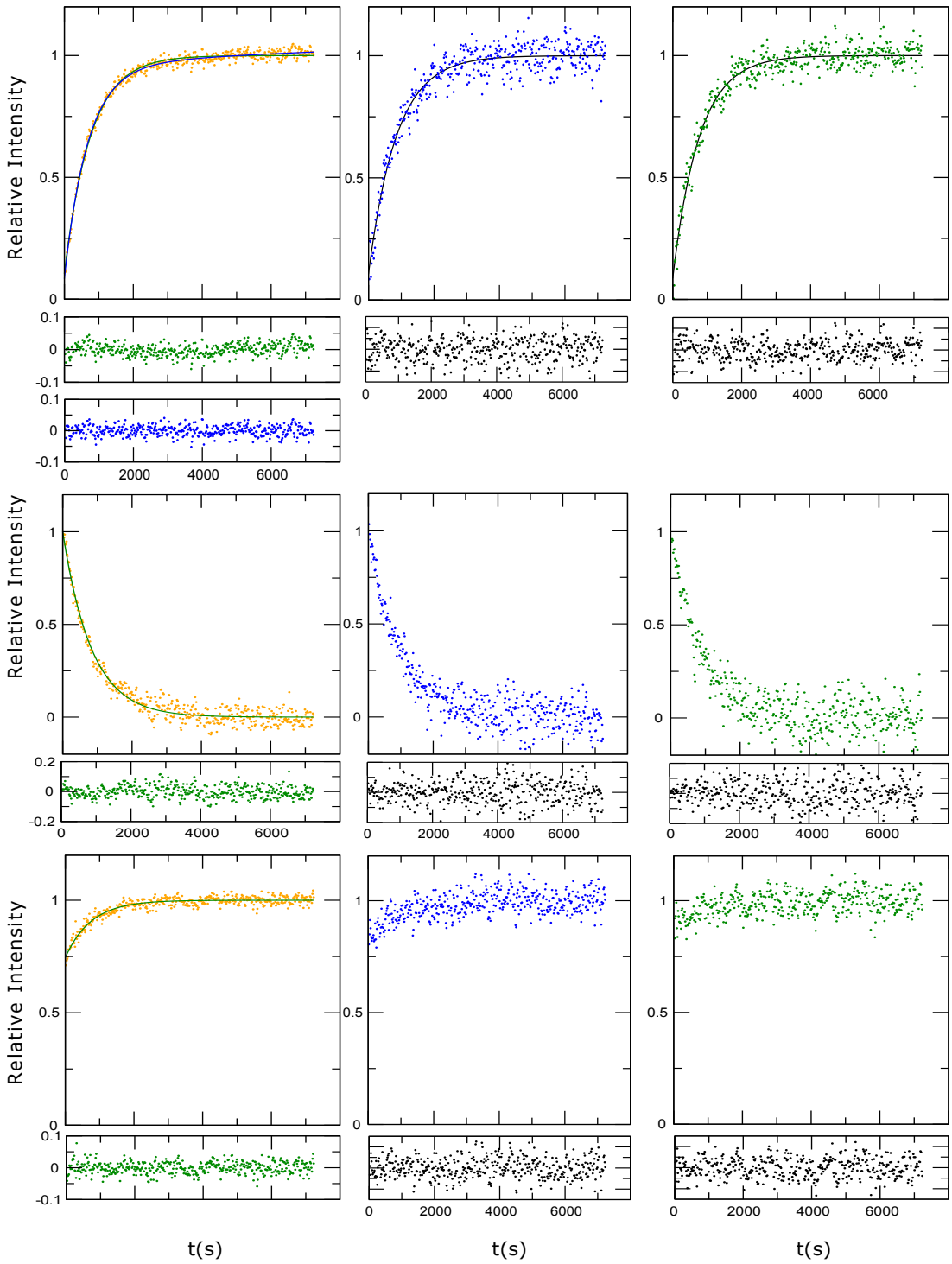


Figure 14.1: Folding of WT-B2M at 25°C. Top line: N class. Center line: I class. Bottom line: I+N class. Left column: 750µM B2M. Center column: 375µM B2M. Right column: 188µM B2M.

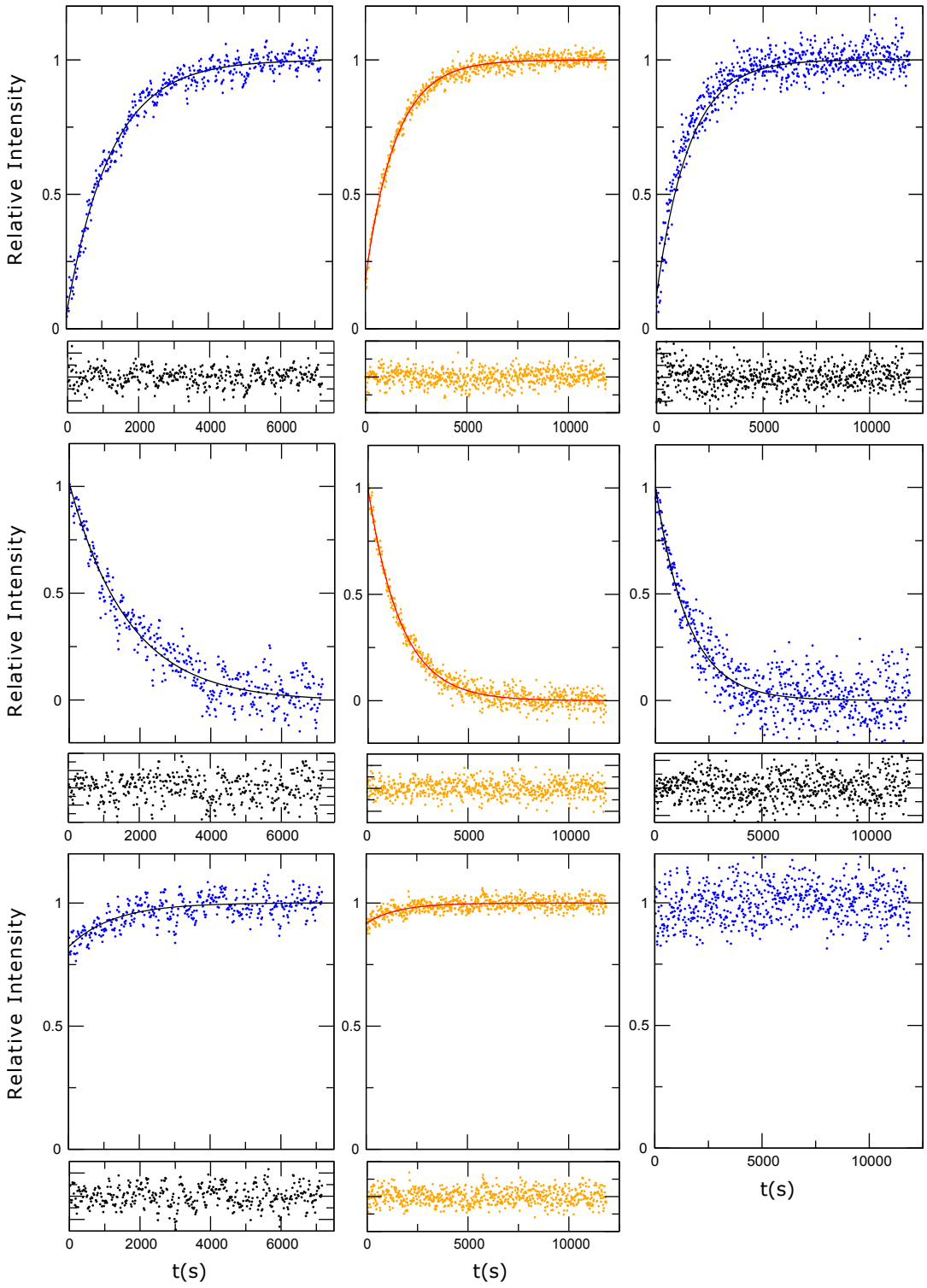


Figure 14.3: Folding of WT-B2M at 20°C. Top line: N class. Center line: I class. Bottom line: I+N class. Left column: 188µM B2M. Center column: 150µM B2M (methyl probes). Right column: 50µM B2M (methyl probes).

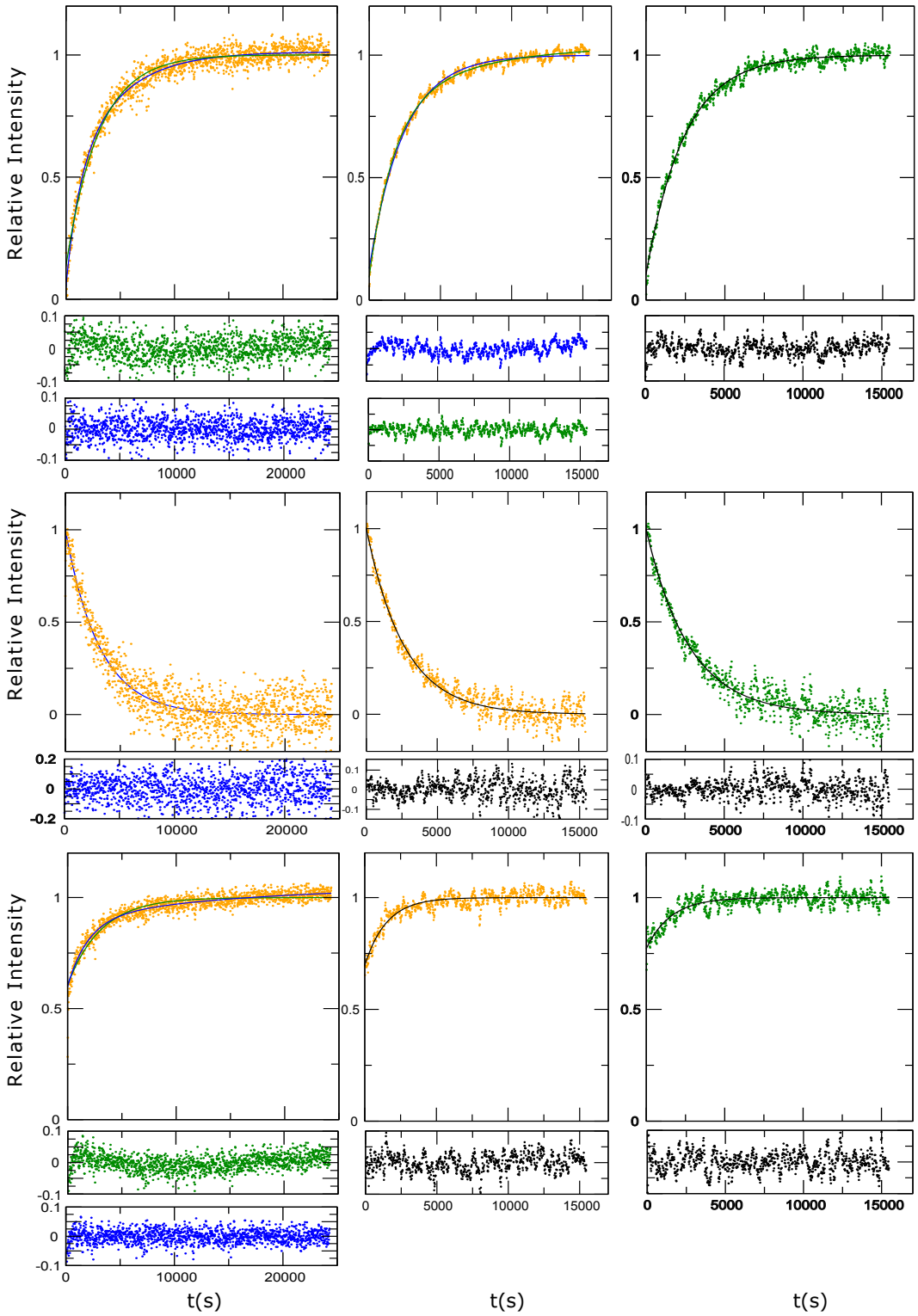


Figure 14.4: Folding of WT-B2M at 15°C. Top line: N class. Center line: I class. Bottom line: I+N class. Left column: 750µM B2M. Center column: 375µM B2M. Right column: 188µM B2M.

Isotopically enriched human B2M production protocol

Buffers

Lysis buffer

10 mM Tris-HCl pH 7.5

2.5 mM MgCl2

0.5 mM CaCl2

0.013 mg/ml DNase I

4 x 10-3 mg/ml lysozyme

10X Triton buffer

50 mM Tris-HCl pH 8.0

100 mM NaCl

0.5% Triton X-100

Solubilisation buffer

8 M Urea

50 mM MES pH 6.5

0.1 mM EDTA

0.1 mM DTT

Refolding Buffer

100 mM Tris-HCl pH 8.0

400 mM L-Arginine-HCl

2 mM EDTA

5 mM Glutathione- Reduced

0.5 mM Glutathione- Oxidised

0.1 mM PMSF

Method

- Preculture one day in LB medium, and then overnight in M9 medium.
- Grow transformed bacteria on M9 media for labelled growth at 37° C
  - Induce at OD600 of ~1.0
  - Harvest at least 12 hrs later
- Harvest bacteria by centrifugation
  - 20 min @ 5000 x g
  - resuspend pellet in lysis buffer (~10ml per gram original cell paste)
- Lyse bacteria by sonification and collect inclusion body (IB) pellet
  - 30 sec on, 30 sec off, 10 min process time in ice bath
  - spin down for 20 min @ 16000 x g
  - Retain pellet, decant supernatant
- Wash IB Pellet in lysis buffer (~10ml per gram original cell paste)
  - homogenise pellet in solution
  - Incubate at 37° C for 1 hour
  - Spin down for 20 min @ 16000 x g
  - Retain pellet, decant supernatant
- Wash IB pellet in triton buffer until pure
  - homogenise in 10x triton buffer (30 ml), spin down for 20 min @ 5000 x g
  - homogenise in 1X triton buffer (30 ml, 1:10 dilution of 10X buffer), spin down for 20 min @ 8000 x g (repeat as necessary until pure (usually two or three times is enough))
  - homogenise in 1X triton buffer (30 ml), spin down for 25 min @ 16000 x g
  - Retain pellet, decant supernatant
- Solubilise IB pellet in solubilisation buffer (~10 ml for each gram of original cell paste)
  - Homogenise and leave with gentle stirring at room temperature overnight
  - Spin down for 20 min @ 16000 x g, retain supernatant
- Flash refolding: Dilute solubilised β2m in 2 litres of refold buffer for one litre of M9 culture

  - Slow dilution (drop by drop) of β2m at 4 ° C with slow agitation

- Concentrate refolded β2m using Vivaspin system and a 5kDa membrane until a volume of ~50mL (~1mg/mL)
- Purify monomer by size exclusion chromatography
  - Sephadex 75 16/60 column (GE Healthcare) used, 5 ml β2m solution loaded per run

Interparticular effects in SAXS: WT B2M pH=2

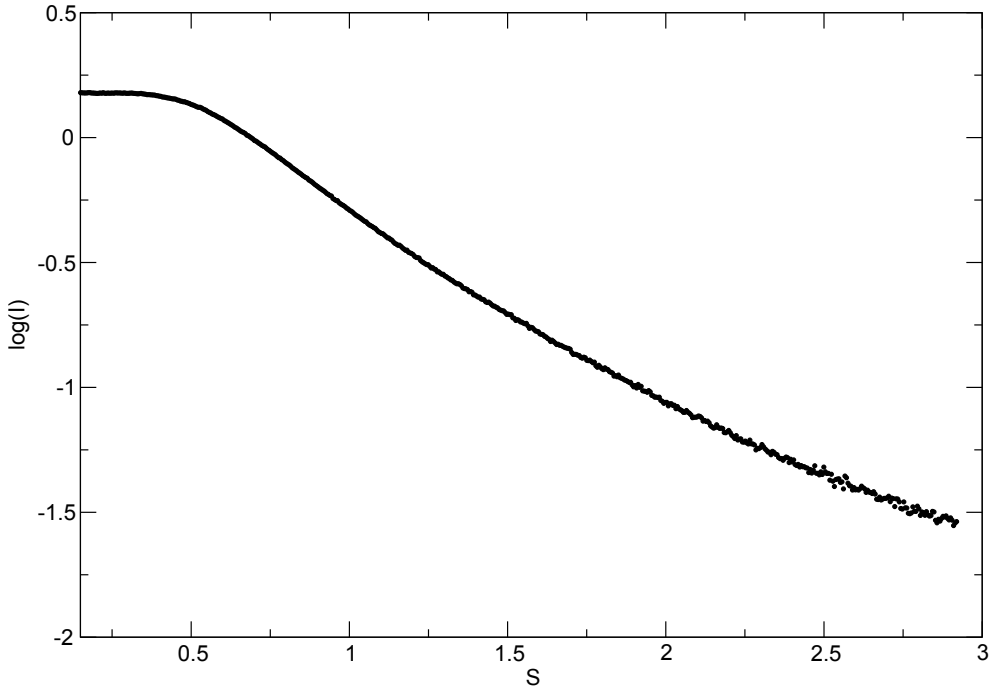


Figure 14.5: SAXS curve WT B2M at pH=2. The plateau at low S is characteristic of repulsive interparticular effects, linked to the acidic denatured state (unfolded and charged) of WT B2M at pH=2.

Mapping of the interaction between the folding intermediate state I_1 and ANS

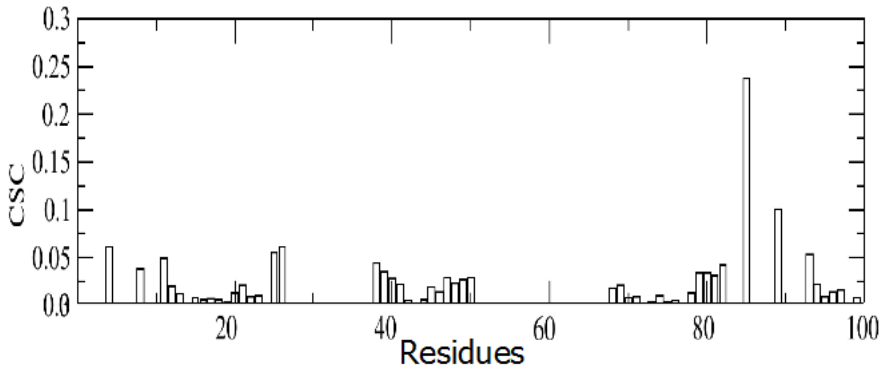


Figure 14.6: Chemical shift changes upon interaction of I_1 and ANS, obtained during the folding of B2M. A binding appears in the 87-93 region of B2M, which corresponds to the apical, opposite to P32, part of the protein.

Bibliography

[1] N. Adachi. Beta-2-microglobulin levels in the cerebrospinal fluid: their value as a disease marker. a review of the recent literature. *Eur Neurol*, 31(4):181–185, 1991.

[2] Cynthia Akkermans, Paul Venema, Salman Rogers, Atze van der Goot, Remko Boom, and Erik van der Linden. Shear pulses nucleate fibril aggregation. *Food Biophysics*, 1:144–150, 2006. 10.1007/s11483-006-9012-5.

[3] C. J. Aldrich, R. Waltrip, E. Hermel, M. Attaya, K. F. Lindahl, J. J. Monaco, and J. Forman. T cell recognition of qa-1b antigens on cells lacking a functional tap-2 transporter. *J Immunol*, 149(12):3773–3777, Dec 1992.

[4] A. Alzheimer, R. A. Stelzmann, H. N. Schnitzlein, and F. R. Murtagh. An english translation of alzheimer's 1907 paper, "uber eine eigenartige erkankung der hirnrinde". *Clin Anat*, 8(6):429–431, 1995.

[5] Carlos Amero, Paul Schanda, M. Asunci?n Dur?, Isabel Ayala, Dominique Marion, Bruno Franzetti, Bernhard Brutscher, and J?r?me Boisbouvier. Fast two-dimensional nmr spectroscopy of high molecular weight protein assemblies. *J Am Chem Soc*, 131(10):3448–3449, Mar 2009.

[6] James M Aramini, Jeffrey L Mills, Rong Xiao, Thomas B Acton, Maggie J Wu, Thomas Szyperski, and Gaetano T Montelione. Resonance assignments for the hypothetical protein yggu from escherichia coli. *J Biomol NMR*, 27(3):285–286, Nov 2003.

[7] Reto Bader, Richard Bamford, Jes?s Zurdo, Ben F Luisi, and Christopher M Dobson. Probing the mechanism of amyloidogenesis through a tandem repeat of the pi3-sh3 domain suggests a generic model for protein aggregation and fibril formation. *J Mol Biol*, 356(1):189–208, Feb 2006.

[8] Emeline Barbet-Massin, Stefano Ricagno, J?zef R Lewandowski, Sofia Giorgetti, Vittorio Bellotti, Martino Bolognesi, Lyndon Emsley, and Guido Pintacuda. Fibrillar vs crystalline full-length beta-2-microglobulin studied by high-

resolution solid-state nmr spectroscopy. *J Am Chem Soc*, 132(16):5556–5557, Apr 2010.

[9] Andreas Barth. Infrared spectroscopy of proteins. *Biochim Biophys Acta*, 1767(9):1073–1101, Sep 2007.

[10] J. W. Becker and G. N. Reeke. Three-dimensional structure of beta 2-microglobulin. *Proc Natl Acad Sci U S A*, 82(12):4225–4229, Jun 1985.

[11] V. Bellotti, M. Stoppini, P. Mangione, M. Sunde, C. Robinson, L. Asti, D. Brancaccio, and G. Ferri. Beta2-microglobulin can be refolded into a native state from ex vivo amyloid fibrils. *Eur J Biochem*, 258(1):61–67, Nov 1998.

[12] N. C. Berchtold and C. W. Cotman. Evolution in the conceptualization of dementia and alzheimer's disease: Greco-roman period to the 1960s. *Neurobiol Aging*, 19(3):173–189, 1998.

[13] I. Bergg?rd and A. G. Bearn. Isolation and properties of a low molecular weight beta-2-globulin occurring in human biological fluids. *J Biol Chem*, 243(15):4095–4103, Aug 1968.

[14] Joanne F Berson, Alexander C Theos, Dawn C Harper, Danielle Tenza, Gra?a Raposo, and Michael S Marks. Proprotein convertase cleavage liberates a fibrillogenic fragment of a resident glycoprotein to initiate melanosome biogenesis. *J Cell Biol*, 161(3):521–533, May 2003.

[15] J. M. Bilbao, E. Horvath, A. R. Hudson, and K. Kovacs. Pituitary adenoma producing amyloid-like substance. *Arch Pathol*, 99(8):411–415, Aug 1975.

[16] G. Bitan, A. Lomakin, and D. B. Teplow. Amyloid beta-protein oligomerization: prenucleation interactions revealed by photo-induced cross-linking of unmodified proteins. *J Biol Chem*, 276(37):35176–35184, Sep 2001.

[17] Gal Bitan, Marina D Kirkitadze, Aleksey Lomakin, Sabrina S Vollers, George B Benedek, and David B Teplow. Amyloid beta -protein (abeta) assembly: Abeta 40 and abeta 42 oligomerize through distinct pathways. *Proc Natl Acad Sci U S A*, 100(1):330–335, Jan 2003.

[18] P. J. Bjorkman, M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger, and D. C. Wiley. Structure of the human class i histocompatibility antigen, hla-a2. *Nature*, 329(6139):506–512, 1987.

[19] Dorottya V Blaho and Andrew D Miranker. Delineating the conformational elements responsible for cu(2+)-induced oligomerization of beta-2 microglobulin. *Biochemistry*, 48(28):6610–6617, Jul 2009.

[20] F. Bloch. Dynamical theory of nuclear induction. ii. *Phys. Rev.*, 102:104–135, Apr 1956.

[21] J. Boffa. Beta-2 microglobulin as a marker for hiv infection. *J Insur Med*, 26(1):14–15, 1994.

[22] L. Bonar, A. S. Cohen, and M. M. Skinner. Characterization of the amyloid fibril as a cross-beta protein. *Proc Soc Exp Biol Med*, 131(4):1373–1375, Sep 1969.

[23] Sylvie Bonin-Guillaume, Dina Zekry, Ezio Giacobini, Gabriel Gold, and Jean-Pierre Michel. [the economical impact of dementia]. *Presse Med*, 34(1):35–41, Jan 2005.

[24] Antoni J H Borysik, Sheena E Radford, and Alison E Ashcroft. Co-populated conformational ensembles of beta2-microglobulin uncovered quantitatively by electrospray ionization mass spectrometry. *J Biol Chem*, 279(26):27069–27077, Jun 2004.

[25] L. F. Boyd, S. Kozlowski, and D. H. Margulies. Solution binding of an antigenic peptide to a major histocompatibility complex class i molecule and the role of beta 2-microglobulin. *Proc Natl Acad Sci U S A*, 89(6):2242–2246, Mar 1992.

[26] Brown and Royer. Fluorescence spectroscopy as a tool to investigate protein interactions. *Curr Opin Biotechnol*, 8(1):45–49, Feb 1997.

[27] Matthew F Calabrese and Andrew D Miranker. Formation of a stable oligomer of beta-2 microglobulin requires only transient encounter with cu(ii). *J Mol Biol*, 367(1):1–7, Mar 2007.

[28] J. M. Campistol, M. Sol?, J. A. Bombi, R. Rodriguez, E. Mirapeix, J. Mu?oz-Gomez, and O. W. Revert. In vitro spontaneous synthesis of beta 2-microglobulin amyloid fibrils in peripheral blood mononuclear cell culture. *Am J Pathol*, 141(1):241–247, Jul 1992.

[29] C. Capeillere-Blandin, T. Delaveau, and B. Descamps-Latscha. Structural modifications of human beta 2 microglobulin treated with oxygen-derived radicals. *Biochem J*, 277 (Pt 1):175–182, Jul 1991.

[30] E. M. Casta?o, J. Ghiso, F. Prelli, P. D. Gorevic, A. Migheli, and B. Frangione. In vitro formation of amyloid fibrils from two synthetic peptides of different lengths homologous to alzheimer's disease beta-protein. *Biochem Biophys Res Commun*, 141(2):782–789, Dec 1986.

[31] Matthew R Chapman, Lloyd S Robinson, Jerome S Pinkner, Robyn Roth, John Heuser, Marten Hammar, Staffan Normark, and Scott J Hultgren. Role of

escherichia coli curli operons in directing amyloid fiber formation. *Science*, 295(5556):851–855, Feb 2002.

[32] S. B. Charg?, E. J. de Koning, and A. Clark. Effect of ph and insulin on fibrillogenesis of islet amyloid polypeptide in vitro. *Biochemistry*, 34(44):14588–14593, Nov 1995.

[33] B. Charra, E. Calemard, M. Uzan, J. C. Terrat, T. Vanel, and G. Laurent. Carpal tunnel syndrome, shoulder pain and amyloid deposits in long-term haemodialysis patients. *Proc Eur Dial Transplant Assoc Eur Ren Assoc*, 21:291–295, 1985.

[34] H. N. Cheng and F. A. Bovey. Cis-trans equilibrium and kinetic studies of acetyl-l-proline and glycyl-l-proline. *Biopolymers*, 16(7):1465–1472, Jul 1977.

[35] Peter Chien, Jonathan S Weissman, and Angela H DePace. Emerging principles of conformation-based prion inheritance. *Annu Rev Biochem*, 73:617–656, 2004.

[36] F. Chiti, E. De Lorenzi, S. Grossi, P. Mangione, S. Giorgetti, G. Caccialanza, C. M. Dobson, G. Merlini, G. Ramponi, and V. Bellotti. A partially structured species of beta 2-microglobulin is significantly populated under physiological conditions and involved in fibrillogenesis. *J Biol Chem*, 276(50):46714–46721, Dec 2001.

[37] F. Chiti, P. Mangione, A. Andreola, S. Giorgetti, M. Stefani, C. M. Dobson, V. Bellotti, and N. Taddei. Detection of two partially structured species in the folding process of the amyloidogenic protein beta 2-microglobulin. *J Mol Biol*, 307(1):379–391, Mar 2001.

[38] Fabrizio Chiti and Christopher M Dobson. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*, 75:333–366, 2006.

[39] A. S. COHEN and E. CALKINS. Electron microscopic observations on a fibrous component in amyloid of diverse origins. *Nature*, 183(4669):1202–1203, Apr 1959.

[40] F. E. Cohen, K. M. Pan, Z. Huang, M. Baldwin, R. J. Fletterick, and S. B. Prusiner. Structural clues to prion replication. *Science*, 264(5158):530–531, Apr 1994.

[41] Fred E Cohen and Jeffery W Kelly. Therapeutic approaches to protein-misfolding diseases. *Nature*, 426(6968):905–909, Dec 2003.

[42] W. Colon and J. W. Kelly. Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry*, 31(36):8654–8660, Sep 1992.

[43] L. H. Connors, T. Shirahama, M. Skinner, A. Fenves, and A. S. Cohen. In vitro formation of amyloid fibrils from intact beta 2-microglobulin. *Biochem Biophys Res Commun*, 131(3):1063–1068, Sep 1985.

[44] C. Cox and T. Lectka. Synthetic catalysis of amide isomerization. *Acc Chem Res*, 33(12):849–858, Dec 2000.

[45] Dalia Daujotyte, Giedrius Vilkaitis, Laura Manelyte, Jack Skalicky, Thomas Szyperski, and Saulius Klimasauskas. Solubility engineering of the hhai methyltransferase. *Protein Eng*, 16(4):295–301, Apr 2003.

[46] A. M. Davison. beta 2-microglobulin and amyloidosis: who is at risk? *Nephrol Dial Transplant*, 10 Suppl 10:48–51, 1995.

[47] S. J. DeArmond, M. P. McKinley, R. A. Barry, M. B. Braunfeld, J. R. McColloch, and S. B. Prusiner. Identification of prion amyloid filaments in scrapie-infected brain. *Cell*, 41(1):221–235, May 1985.

[48] Christopher M. Dobson, Andrej ? ali, and Martin Karplus. Protein folding: A perspective from theory and experiment. *Angewandte Chemie International Edition*, 37(7):868–893, 1998.

[49] Katarzyna Domanska, Saskia Vanderhaegen, Vasundara Srinivasan, Els Pardon, Florine Dupeux, Jose A Marquez, Sofia Giorgetti, Monica Stoppini, Lode Wyns, Vittorio Bellotti, and Jan Steyaert. Atomic structure of a nanobody-trapped domain-swapped dimer of an amyloidogenic beta2-microglobulin variant. *Proc Natl Acad Sci U S A*, 108(4):1314–1319, Jan 2011.

[50] Mireille Dumoulina and Reto Bader. A short historical survey of developments in amyloid research. *Protein Pept Lett*, 13(3):213–217, 2006.

[51] Nicholas F Dupuis, Chun Wu, Joan-Emma Shea, and Michael T Bowers. The amyloid formation mechanism in human iapp: dimers have ?-strand monomer-monomer interfaces. *J Am Chem Soc*, 133(19):7240–7243, May 2011.

[52] Catherine M Eakin, Jefferson D Knight, Charles J Morgan, Michael A Gelfand, and Andrew D Miranker. Formation of a copper specific binding site in non-native states of beta-2-microglobulin. *Biochemistry*, 41(34):10646–10656, Aug 2002.

[53] E. D. Eanes and G. G. Glenner. X-ray diffraction studies on amyloid filaments. *J Histochem Cytochem*, 16(11):673–677, Nov 1968.

[54] Timo Eichner, Arnout P Kalverda, Gary S Thompson, Steve W Homans, and Sheena E Radford. Conformational conversion during amyloid formation at atomic resolution. *Mol Cell*, 41(2):161–172, Jan 2011.

[55] Timo Eichner and Sheena E Radford. Understanding the complex mechanisms of ?(2) -microglobulin amyloid assembly. *FEBS J*, May 2011.

[56] T. Elliott, J. Elvin, V. Cerundolo, H. Allen, and A. Townsend. Structural requirements for the peptide-induced conformational change of free major histocompatibility complex class i heavy chains. *Eur J Immunol*, 22(8):2085–2091, Aug 1992.

[57] RR Ernst and WA Anderson. Application of fourier transform spectroscopy to magnetic resonance. *Review of Scientific Instruments*, 37:93, 1966.

[58] G. Esposito, R. Michelutti, G. Verdone, P. Viglino, H. Hern?ndez, C. V. Robinson, A. Amoresano, F. Dal Piaz, M. Monti, P. Pucci, P. Mangione, M. Stoppini, G. Merlini, G. Ferri, and V. Bellotti. Removal of the n-terminal hexapeptide from human beta2-microglobulin facilitates protein aggregation and fibril formation. *Protein Sci*, 9(5):831–845, May 2000.

[59] Gennaro Esposito, Stefano Ricagno, Alessandra Corazza, Enrico Rennella, Devrim G?mral, Maria Chiara Mimmi, Elena Betto, Carlo E M Pucillo, Federico Fogolari, Paolo Viglino, Sara Raimondi, Sofia Giorgetti, Benedetta Bolognesi, Giampaolo Merlini, Monica Stoppini, Martino Bolognesi, and Vittorio Bellotti. The controlling roles of trp60 and trp95 in beta2-microglobulin function, folding and amyloid aggregation properties. *J Mol Biol*, 378(4):887–897, May 2008.

[60] Adrien Favier and Bernhard Brutscher. Recovering lost magnetization: polarization enhancement in biomolecular nmr. *J Biomol NMR*, 49(1):9–15, Jan 2011.

[61] Jocelyne Fiaux, Z. Petek Cakar, Marco Sonderegger, Kurt W?thrich, Thomas Szyperski, and Uwe Sauer. Metabolic-flux profiling of the yeasts saccharomyces cerevisiae and pichia stipitis. *Eukaryot Cell*, 2(1):170–180, Feb 2003.

[62] M. F?ndrich, M. A. Fletcher, and C. M. Dobson. Amyloid fibrils from muscle myoglobin. *Nature*, 410(6825):165–166, Mar 2001.

[63] Douglas M Fowler, Atanas V Koulov, Christelle Alory-Jost, Michael S Marks, William E Balch, and Jeffery W Kelly. Functional amyloid formation within mammalian tissue. *PLoS Biol*, 4(1):e6, Jan 2006.

[64] F. Gejyo, N. Homma, Y. Suzuki, and M. Arakawa. Serum levels of beta 2-microglobulin as a new form of amyloid protein in patients undergoing long-term hemodialysis. *N Engl J Med*, 314(9):585–586, Feb 1986.

[65] F. Gejyo, T. Yamada, S. Odani, Y. Nakagawa, M. Arakawa, T. Kunitomo, H. Kataoka, M. Suzuki, Y. Hirasawa, and T. Shirahama. A new form of amyloid protein associated with chronic hemodialysis was identified as beta 2-microglobulin. *Biochem Biophys Res Commun*, 129(3):701–706, Jun 1985.

[66] Morie A Gertz. The classification and typing of amyloid deposits. *Am J Clin Pathol*, 121(6):787–789, Jun 2004.

[67] C. Geula, C. K. Wu, D. Saroff, A. Lorenzo, M. Yuan, and B. A. Yankner. Aging renders the brain vulnerable to amyloid beta-protein neurotoxicity. *Nat Med*, 4(7):827–831, Jul 1998.

[68] Sofia Giorgetti, Antonio Rossi, Palma Mangione, Sara Raimondi, Sara Marini, Monica Stoppini, Alessandra Corazza, Paolo Viglino, Gennaro Esposito, Giuseppe Cetta, Giampaolo Merlini, and Vittorio Bellotti. Beta2-microglobulin isoforms display an heterogeneous affinity for type i collagen. *Protein Sci*, 14(3):696–702, Mar 2005.

[69] G. G. Glenner. Amyloid deposits and amyloidosis. the beta-fibrilloses (first of two parts). *N Engl J Med*, 302(23):1283–1292, Jun 1980.

[70] G. G. Glenner. Amyloid deposits and amyloidosis: the beta-fibrilloses (second of two parts). *N Engl J Med*, 302(24):1333–1343, Jun 1980.

[71] G. G. Glenner, E. D. Eanes, H. A. Bladen, R. P. Linke, and J. D. Termine. Beta-pleated sheet fibrils. a comparison of native amyloid with synthetic protein fibrils. *J Histochem Cytochem*, 22(12):1141–1158, Dec 1974.

[72] G. G. Glenner, D. Ein, E. D. Eanes, H. A. Bladen, W. Terry, and D. L. Page. Creation of "amyloid" fibrils from bence jones proteins in vitro. *Science*, 174(10):712–714, Nov 1971.

[73] G. G. Glenner and D. L. Page. Amyloid, amyloidosis, and amyloidogenesis. *Int Rev Exp Pathol*, 15:1–92, 1976.

[74] L. G. Goldfarb, P. Brown, M. Haltia, J. Ghiso, B. Frangione, and D. C. Gajdusek. Synthetic peptides corresponding to different mutated regions of the amyloid gene in familial creutzfeldt-jakob disease show enhanced in vitro formation of morphologically different amyloid fibrils. *Proc Natl Acad Sci U S A*, 90(10):4451–4454, May 1993.

[75] P. D. Gorevic, T. T. Casey, W. J. Stone, C. R. DiRaimondo, F. C. Prelli, and B. Frangione. Beta-2 microglobulin is an amyloidogenic protein in man. *J Clin Invest*, 76(6):2425–2429, Dec 1985.

[76] H. M. Grey, R. T. Kubo, S. M. Colon, M. D. Poulik, P. Cresswell, T. Springer, M. Turner, and J. L. Strominger. The small subunit of hl-a antigens is beta 2-microglobulin. *J Exp Med*, 138(6):1608–1612, Dec 1973.

[77] Yvonne Groemping and Nadja Hellmann. Spectroscopic methods for the determination of protein interactions. *Curr Protoc Protein Sci*, Chapter 20:Unit 20.8, Mar 2005.

[78] M. Gross, D. K. Wilkins, M. C. Pitkeathly, E. W. Chung, C. Higham, A. Clark, and C. M. Dobson. Formation of amyloid fibrils by peptides derived from the bacterial cold shock protein cspb. *Protein Sci*, 8(6):1350–1357, Jun 1999.

[79] J. I. Guijarro, M. Sunde, J. A. Jones, I. D. Campbell, and C. M. Dobson. Amyloid fibril formation by an sh3 domain. *Proc Natl Acad Sci U S A*, 95(8):4224–4228, Apr 1998.

[80] J. Hardy and D. Allsop. Amyloid deposition as the central event in the aetiology of alzheimer's disease. *Trends Pharmacol Sci*, 12(10):383–388, Oct 1991.

[81] J. D. Harper, C. M. Lieber, and P. T. Lansbury. Atomic force microscopic imaging of seeded fibril formation and fibril branching by the alzheimer's disease amyloid-beta protein. *Chem Biol*, 4(12):951–959, Dec 1997.

[82] Elizabeth Head and Ira T Lott. Down syndrome and beta-amyloid deposition. *Curr Opin Neurol*, 17(2):95–100, Apr 2004.

[83] Niels H H Heegaard, Peter Roepstorff, Steen G Melberg, and Mogens H Nissen. Cleaved beta 2-microglobulin partially attains a conformation that has amyloidogenic features. *J Biol Chem*, 277(13):11184–11189, Mar 2002.

[84] Hagen Hofmann, Frank Hillger, Shawn H Pfeil, Armin Hoffmann, Daniel Streich, Dominik Haenni, Daniel Nettels, Everett A Lipman, and Benjamin Schuler. Single-molecule spectroscopy of protein folding in a chaperonin cage. *Proc Natl Acad Sci U S A*, 107(26):11793–11798, Jun 2010.

[85] Philip J Hogg. Disulfide bonds as switches for protein function. *Trends Biochem Sci*, 28(4):210–214, Apr 2003.

[86] Qingling Huang, Linlin Ding, Kim B Phan, Catherine Cheng, Chun hong Xia, Xiaohua Gong, and Joseph Horwitz. Mechanism of cataract formation in alphaa-crystallin y118d mutation. *Invest Ophthalmol Vis Sci*, 50(6):2919–2926, Jun 2009.

[87] Ling Chin Hwang and Thorsten Wohland. Recent advances in fluorescence cross-correlation spectroscopy. *Cell Biochem Biophys*, 49(1):1–13, 2007.

[88] F. Hyafil and J. L. Strominger. Dissociation and exchange of the beta 2-microglobulin subunit of hla-a and hla-b antigens. *Proc Natl Acad Sci U S A*, 76(11):5834–5838, Nov 1979.

[89] D. E. Isenman, R. H. Painter, and K. J. Dorrington. The structure and function of immunoglobulin domains: studies with beta-2-microglobulin on the role of the intrachain disulfide bond. *Proc Natl Acad Sci U S A*, 72(2):548–552, Feb 1975.

[90] Andreas Jabs. Determination of secondary structure in proteins by fourier transform infrared spectroscopy (ftir). *Jena Library of Biological Macromolecules*.

[91] M. Jadoul, C. Garbar, H. No?l, J. Sennesael, R. Vanholder, P. Bernaert, G. Rorive, G. Hanique, and C. van Ypersele de Strihou. Histological prevalence of beta 2-microglobulin amyloidosis in hemodialysis: a prospective post-mortem study. *Kidney Int*, 51(6):1928–1932, Jun 1997.

[92] Thomas R Jahn, Martin J Parker, Steve W Homans, and Sheena E Radford. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat Struct Mol Biol*, 13(3):195–201, Mar 2006.

[93] Thomas R Jahn and Sheena E Radford. Folding versus aggregation: polypeptide conformations on competing pathways. *Arch Biochem Biophys*, 469(1):100–117, Jan 2008.

[94] T. S. Jardetzky, W. S. Lane, R. A. Robinson, D. R. Madden, and D. C. Wiley. Identification of self peptides bound to purified hla-b27. *Nature*, 353(6342):326–329, Sep 1991.

[95] Jos? L Jim?nez, Ewan J Nettleton, Mario Bouchard, Carol V Robinson, Christopher M Dobson, and Helen R Saibil. The protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci U S A*, 99(14):9196–9201, Jul 2002.

[96] Lee-Way Jin, Kacey A. Claborn, Miki Kurimoto, Morten A. Geday, Izumi Maezawa, Faranak Sohraby, Marcus Estrada, Werner Kaminksy, and Bart Kahr. Imaging linear birefringence and dichroism in cerebral amyloid pathologies. *Proceedings of the National Academy of Sciences*, 100(26):15294–15298, 2003.

[97] Atsushi Kameda, Masaru Hoshino, Takashi Higurashi, Satoshi Takahashi, Hironobu Naiki, and Yuji Goto. Nuclear magnetic resonance characterization of the refolding intermediate of beta2-microglobulin trapped by non-native prolyl peptide bond. *J Mol Biol*, 348(2):383–397, Apr 2005.

[98] Rakez Kayed, Elizabeth Head, Jennifer L Thompson, Theresa M McIntire, Saskia C Milton, Carl W Cotman, and Charles G Glabe. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science*, 300(5618):486–489, Apr 2003.

[99] I. Kedar, M. Ravid, and E. Sohar. In vitro synthesis of "amyloid"fibrils from insulin, calcitonin and parathormone. *Isr J Med Sci*, 12(10):1137–1140, Oct 1976.

[100] Alexander E Kelly, Horng D Ou, Richard Withers, and Volker D?tsch. Low-conductivity buffers for high-sensitivity nmr measurements. *J Am Chem Soc*, 124(40):12013–12019, Oct 2002.

[101] Sharon M Kelly and Nicholas C Price. Circular dichroism to study protein interactions. *Curr Protoc Protein Sci*, Chapter 20:Unit 20.10, Dec 2006.

[102] Seho Kim and Thomas Szyperski. Gft nmr, a new approach to rapidly obtain precise high-dimensional nmr spectral information. *J Am Chem Soc*, 125(5):1385–1393, Feb 2003.

[103] Ian R. Kleckner and Mark P. Foster. An introduction to nmr-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics*, 1814(8):942 – 968, 2011. <ce:title>Protein Dynamics: Experimental and Computational Approaches</ce:title>.

[104] S. Kozlowski, T. Takeshita, W. H. Boehncke, H. Takahashi, L. F. Boyd, R. N. Germain, J. A. Berzofsky, and D. H. Margulies. Excess beta 2 microglobulin promoting functional peptide association with purified soluble class i mhc molecules. *Nature*, 349(6304):74–77, Jan 1991.

[105] E. Kupce, J. Boyd, and I. D. Campbell. Short selective pulses for biochemical applications. *J Magn Reson B*, 106(3):300–303, Mar 1995.

[106] J. K. Lee, S. C. Tsai, J. F. Hsieh, Y. J. Ho, S. S. Sun, and C. H. Kao. Beta-2-microglobulin (beta 2m) as a tumor marker in nasopharyngeal carcinoma. *Anticancer Res*, 20(6C):4765–4768, 2000.

[107] R. P. Linke, F. W. Tischendorf, D. Zucker-Franklin, and E. C. Franklin. The formation of amyloid-like fibrils in vitro from bence jones proteins of the vlambdai subclass. *J Immunol*, 111(1):24–26, Jul 1973.

[108] Cong Liu, Michael R Sawaya, and David Eisenberg. ??-microglobulin forms three-dimensional domain-swapped amyloid fibrils with disulfide linkages. *Nat Struct Mol Biol*, 18(1):49–55, Jan 2011.

[109] Gaohua Liu, Jeffrey L Mills, Tracy A Hess, Seho Kim, Jack J Skalicky, Dinesh K Sukumaran, Eriks Kupce, Arne Skerra, and Thomas Szyperski. Resonance assignments for the 21 kda engineered fluorescein-binding lipocalin flua. *J Biomol NMR*, 27(2):187–188, Oct 2003.

[110] A. Lorenzo, B. Razzaboni, G. C. Weir, and B. A. Yankner. Pancreatic islet cell toxicity of amylin associated with type-2 diabetes mellitus. *Nature*, 368(6473):756–760, Apr 1994.

[111] L. F. Lue, Y. M. Kuo, A. E. Roher, L. Brachova, Y. Shen, L. Sue, T. Beach, J. H. Kurth, R. E. Rydel, and J. Rogers. Soluble amyloid beta peptide concentration as a predictor of synaptic change in alzheimer's disease. *Am J Pathol*, 155(3):853–862, Sep 1999.

[112] D. R. Madden, J. C. Gorga, J. L. Strominger, and D. C. Wiley. The three-dimensional structure of hla-b27 at 2.1 a resolution suggests a general mechanism for tight peptide binding to mhc. *Cell*, 70(6):1035–1048, Sep 1992.

[113] D. Marion, M. Ikura, R. Tschudin, and A. Bax. Rapid recording of 2d nmr spectra without phase cycling. application to the study of hydrogen exchange in proteins. *J. Magn. Reson*, 85(2):393–399, 1989.

[114] Stephen R Martin and Maria J Schilstra. Circular dichroism and its application to the study of biomolecules. *Methods Cell Biol*, 84:263–293, 2008.

[115] J. Masel, V. A. Jansen, and M. A. Nowak. Quantifying the kinetic parameters of prion replication. *Biophys Chem*, 77(2-3):139–152, Mar 1999.

[116] V. J. McParland, N. M. Kad, A. P. Kalverda, A. Brown, P. Kirwin-Jones, M. G. Hunter, M. Sunde, and S. E. Radford. Partially unfolded states of beta(2)-microglobulin and amyloid formation in vitro. *Biochemistry*, 39(30):8735–8746, Aug 2000.

[117] D. Moechars, I. Dewachter, K. Lorent, D. Revers?, V. Baekelandt, A. Naidu, I. Tesseur, K. Spittaels, C. V. Haute, F. Checler, E. Godaux, B. Cordell, and F. Van Leuven. Early phenotypic changes in transgenic mice that overexpress different mutants of amyloid precursor protein in brain. *J Biol Chem*, 274(10):6483–6492, Mar 1999.

[118] C. J. Morgan, M. Gelfand, C. Atreya, and A. D. Miranker. Kidney dialysis-associated amyloidosis: a molecular role for copper in fiber formation. *J Mol Biol*, 309(2):339–345, Jun 2001.

[119] Amritpal Mudher and Simon Lovestone. Alzheimer's disease-do tauists and baptists finally shake hands? *Trends Neurosci*, 25(1):22–26, Jan 2002.

[120] Sarah L Myers, Susan Jones, Thomas R Jahn, Isobel J Morten, Glenys A Tennent, Eric W Hewitt, and Sheena E Radford. A systematic study of the effect of physiological factors on beta2-microglobulin amyloid formation at neutral ph. *Biochemistry*, 45(7):2311–2321, Feb 2006.

[121] Hironobu Naiki, Norikazu Hashimoto, Satoru Suzuki, Hideki Kimura, Kazuya Nakakuki, and Fumitake Gejyo. Establishment of a kinetic model of dialysis-related amyloid fibril extension in vitro. *Amyloid*, 4(4):223–232, 1997.

[122] C. Nilsberth, A. Westlind-Danielsson, C. B. Eckman, M. M. Condron, K. Axelman, C. Forsell, C. Stenh, J. Luthman, D. B. Teplow, S. G. Younkin, J. N?slund, and L. Lannfelt. The 'arctic' app mutation (e693g) causes alzheimer's disease by enhanced abeta protofibril formation. *Nat Neurosci*, 4(9):887–893, Sep 2001.

[123] Yumiko Ohhashi, Yoshihisa Hagihara, Gennady Kozhukh, Masaru Hoshino, Kazuhiro Hasegawa, Itaru Yamaguchi, Hironobu Naiki, and Yuji Goto. The intrachain disulfide bond of beta(2)-microglobulin is not essential for the immunoglobulin fold at neutral ph, but is essential for amyloid fibril formation at acidic ph. *J Biochem*, 131(1):45–52, Jan 2002.

[124] S. B. Padrick and A. D. Miranker. Islet amyloid polypeptide: identification of long-range contacts and local order on the fibrillogenesis pathway. *J Mol Biol*, 308(4):783–794, May 2001.

[125] J. Parkinson. *An essay on the shaking palsy.* Printed by Whittingham and Rowland, for Sherwood, Neely, and Jones, 1817.

[126] E. K. Perry, A. E. Oakley, J. M. Candy, and R. H. Perry. Properties and possible significance of substance p and insulin fibrils. *Neurosci Lett*, 25(3):321–325, Sep 1981.

[127] T. A. Pertinhez, M. Bouchard, E. J. Tomlinson, R. Wain, S. J. Ferguson, C. M. Dobson, and L. J. Smith. Amyloid fibril formation by a helical cytochrome. *FEBS Lett*, 495(3):184–186, Apr 2001.

[128] P. A. Peterson, B. A. Cunningham, I. Bergg?rd, and G. M. Edelman. 2 - microglobulin–a free immunoglobulin domain. *Proc Natl Acad Sci U S A*, 69(7):1697–1701, Jul 1972.

[129] Henriett P?l-G?bor, Linda Gombos, Andr?s Micsonai, Erika Kov?cs, Eva Petrik, J?nos Kov?cs, L?szl? Gr?f, Judit Fidy, Hironobu Naiki, Yuji Goto, K?roly Liliom, and J?zsef Kardos. Mechanism of lysophosphatidic acid-induced amyloid fibril formation of beta(2)-microglobulin in vitro under physiological conditions. *Biochemistry*, 48(24):5689–5699, Jun 2009.

[130] M. Pras, M. Schubert, D. Zucker-Franklin, A. Rimon, and E. C. Franklin. The characterization of soluble amyloid prepared in water. *J Clin Invest*, 47(4):924–933, Apr 1968.

[131] S. B. Prusiner. Novel proteinaceous infectious particles cause scrapie. *Science*, 216(4542):136–144, Apr 1982.

[132] M. Ramirez-Alvarado, J. S. Merkel, and L. Regan. A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. *Proc Natl Acad Sci U S A*, 97(16):8979–8984, Aug 2000.

[133] A. G. Redfield. On the theory of relaxation processes. *IBM Journal of Research and Development*, 1, 1957.

[134] A. G. Redfield. *Adv. Magn. Reson.*, 1:1, 1965.

[135] Enrico Rennella, Alessandra Corazza, Sofia Giorgetti, Federico Fogolari, Paolo Viglino, Riccardo Porcari, Laura Verga, Monica Stoppini, Vittorio Bellotti, and Gennaro Esposito. Folding and fibrillogenesis: clues from beta2-microglobulin. *J Mol Biol*, 401(2):286–297, Aug 2010.

[136] G. ROMHANYI. [not available]. *Schweiz Z Pathol Bakteriol*, 12(3):253–262, 1949.

[137] C. Rosano, S. Zuccotti, P. Mangione, S. Giorgetti, V. Bellotti, F. Pettirossi, A. Corazza, P. Viglino, G. Esposito, and M. Bolognesi. beta2-microglobulin h31y variant 3d structure highlights the protein natural propensity towards intermolecular aggregation. *J Mol Biol*, 335(4):1051–1064, Jan 2004.

[138] Camillo Rosano, Simone Zuccotti, and Martino Bolognesi. The three-dimensional structure of beta2 microglobulin: results from x-ray crystallography. *Biochim Biophys Acta*, 1753(1):85–91, Nov 2005.

[139] J. B. Rothbard and M. L. Gefter. Interactions between immunogenic peptides and mhc proteins. *Annu Rev Immunol*, 9:527–565, 1991.

[140] Rahul Roy, Sungchul Hohng, and Taekjip Ha. A practical guide to single-molecule fret. *Nat Methods*, 5(6):507–516, Jun 2008.

[141] Yaroslav E Ryabov, Charles Geraghty, Amitabh Varshney, and David Fushman. An efficient computational method for predicting rotational diffusion tensors of globular proteins using an ellipsoid representation. *J Am Chem Soc*, 128(48):15432–15444, Dec 2006.

[142] Carsten Sachse, Nikolaus Grigorieff, and Marcus F?ndrich. Nanoscale flexibility parameters of alzheimer amyloid fibrils determined by electron cryo-microscopy. *Angew Chem Int Ed Engl*, 49(7):1321–1323, Feb 2010.

[143] Michiko Sakata, Eri Chatani, Atsushi Kameda, Kazumasa Sakurai, Hironobu Naiki, and Yuji Goto. Kinetic coupling of folding and prolyl isomerization of beta2-microglobulin studied by mutational analysis. *J Mol Biol*, 382(5):1242–1255, Oct 2008.

[144] Carlo Santambrogio, Stefano Ricagno, Matteo Colombo, Alberto Barbiroli, Francesco Bonomi, Vittorio Bellotti, Martino Bolognesi, and Rita Grandori. De-loop mutations affect beta2 microglobulin stability, oligomerization, and the low-ph unfolded form. *Protein Sci*, 19(7):1386–1394, Jul 2010.

[145] M. A. Saper, P. J. Bjorkman, and D. C. Wiley. Refined structure of the human histocompatibility antigen hla-a2 at 2.6 a resolution. *J Mol Biol*, 219(2):277–319, May 1991.

[146] Paul Schanda, Vincent Forge, and Bernhard Brutscher. Het-sofast nmr for fast detection of structural compactness and heterogeneity along polypeptide chains. *Magn Reson Chem*, 44 Spec No:S177–S184, Jul 2006.

[147] Paul Schanda, Vincent Forge, and Bernhard Brutscher. Protein folding and unfolding studied at atomic resolution by fast two-dimensional nmr spectroscopy. *Proc Natl Acad Sci U S A*, 104(27):11257–11262, Jul 2007.

[148] Paul Schanda, Eriks Kupce, and Bernhard Brutscher. Sofast-hmqc experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR*, 33(4):199–211, Dec 2005.

[149] S. Schwarzinger, G. J. Kroon, T. R. Foss, J. Chung, P. E. Wright, and H. J. Dyson. Sequence-dependent correction of random coil nmr chemical shifts. *J Am Chem Soc*, 123(13):2970–2978, Apr 2001.

[150] Philipp Selenko, Dominique P Frueh, Simon J Elsaesser, Wilhelm Haas, Steven P Gygi, and Gerhard Wagner. In situ observation of protein phosphorylation by high-resolution nmr spectroscopy. *Nat Struct Mol Biol*, 15(3):321–329, Mar 2008.

[151] T. R. Serio, A. G. Cashikar, A. S. Kowal, G. J. Sawicki, J. J. Moslehi, L. Serpell, M. F. Arnsdorf, and S. L. Lindquist. Nucleated conformational conversion and the replication of conformational information by a prion determinant. *Science*, 289(5483):1317–1321, Aug 2000.

[152] T. Shirahama and A. S. Cohen. High-resolution electron microscopic analysis of the amyloid fibril. *J Cell Biol*, 33(3):679–708, Jun 1967.

[153] Kausik Si, Maurizio Giustetto, Amit Etkin, Ruby Hsu, Agnieszka M Janisiewicz, Maria Conchetta Miniaci, Joung-Hun Kim, Huixiang Zhu, and Eric R Kandel. A neuronal isoform of cpeb regulates local protein synthesis and stabilizes synapse-specific long-term facilitation in aplysia. *Cell*, 115(7):893–904, Dec 2003.

[154] Michael H Simonian and John A Smith. Spectrophotometric and colorimetric determination of protein concentration. *Curr Protoc Mol Biol*, Chapter 10:Unit 10.1A, Nov 2006.

[155] J. P. Simorre, B. Brutscher, M. S. Caffrey, and D. Marion. Assignment of nmr spectra of proteins using triple-resonance two-dimensional experiments. *J Biomol NMR*, 4(3):325–333, May 1994.

[156] J. D. Sipe. Amyloidosis. *Annu Rev Biochem*, 61:947–975, 1992.

[157] J. D. Sipe and A. S. Cohen. Review: history of the amyloid fibril. *J Struct Biol*, 130(2-3):88–98, Jun 2000.

[158] J.D. Sipe. *Amyloid proteins: the beta sheet conformation and disease*. Number vol. 1 in Amyloid Proteins: The Beta Sheet Conformation and Disease. Wiley-VCH, 2005.

[159] Andrew M Smith, Thomas R Jahn, Alison E Ashcroft, and Sheena E Radford. Direct observation of oligomeric species formed in the early stages of amyloid fibril formation using electrospray ionisation mass spectrometry. *J Mol Biol*, 364(1):9–19, Nov 2006.

[160] G. D. Sorenson, W. A. Heefner, and J. B. Kirkpatrick. Experimental amyloidosis. ii. light and electron microscopic observations of liver. *Am J Pathol*, 44(4):629–644, Apr 1964.

[161] D.J. States, R.A. Haberkorn, and D.J. Ruben. A two-dimensional nuclear overhauser experiment with pure absorption phase in four quadrants. *J Magn Reson*, 48:286–292, 1982.

[162] Dmitri I Svergun and Michel H J Koch. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics*, 66(10):1735, 2003.

[163] T. Szyperski, G. Wider, J. H. Bushweller, and K. Wuethrich. Reduced dimensionality in triple-resonance nmr experiments. *Journal of the American Chemical Society*, 115(20):9307–9308, 1993.

[164] Saeed Hesami Takallu, Mostafa Rezaei Tavirani, Shiva Kalantari, Mahrooz Amir Bakhtiarvan, and Sayed Mohammad Mahdavi. Co-amoxiclav effects on the structural and binding properties of human serum albumin. *Iranian Journal of Pharmaceutical Research*, 9:251–257, 2010.

[165] J. D. Termine, E. D. Eanes, D. Ein, and G. G. Glenner. Infrared spectroscopy of human amyloid fibrils and immunoglobulin proteins. *Biopolymers*, 11(5):1103–1113, 1972.

[166] Carl A Thompson, Karen Spilsbury, Jill Hall, Yvonne Birks, Colin Barnes, and Joy Adamson. Systematic review of information and support interventions for caregivers of people with dementia. *BMC Geriatr*, 7:18, 2007.

[167] A. Townsend, C. Ohl?n, J. Bastin, H. G. Ljunggren, L. Foster, and K. K?rre. Association of class i major histocompatibility heavy and light chains induced by viral peptides. *Nature*, 340(6233):443–448, Aug 1989.

[168] Chi H Trinh, David P Smith, Arnout P Kalverda, Simon E V Phillips, and Sheena E Radford. Crystal structure of monomeric human beta-2-microglobulin reveals clues to its amyloidogenic properties. *Proc Natl Acad Sci U S A*, 99(15):9771–9776, Jul 2002.

[169] C. van Ypersele de Strihou, J. Floege, M. Jadoul, and K. M. Koch. Amyloidosis and its relationship to different dialysers. *Nephrol Dial Transplant*, 9 Suppl 2:156–161, 1994.

[170] Giuliana Verdone, Alessandra Corazza, Paolo Viglino, Fabio Pettirossi, Sofia Giorgetti, Palma Mangione, Alessia Andreola, Monica Stoppini, Vittorio Bellotti, and Gennaro Esposito. The solution structure of human beta2-microglobulin reveals the prodromes of its amyloid transition. *Protein Sci*, 11(3):487–499, Mar 2002.

[171] Bente Vestergaard, Minna Groenning, Manfred Roessle, Jette S Kastrup, Marco van de Weert, James M Flink, Sven Frokjaer, Michael Gajhede, and Dmitri I Svergun. A helical structural nucleus is the primary elongating unit of insulin amyloid fibrils. *PLoS Biol*, 5(5):e134, May 2007.

[172] A. Vitiello, T. A. Potter, and L. A. Sherman. The role of beta 2-microglobulin in peptide binding by class i molecules. *Science*, 250(4986):1423–1426, Dec 1990.

[173] D. M. Walsh, D. M. Hartley, Y. Kusumoto, Y. Fezoui, M. M. Condron, A. Lomakin, G. B. Benedek, D. J. Selkoe, and D. B. Teplow. Amyloid beta-protein fibrillogenesis. structure and biological activity of protofibrillar intermediates. *J Biol Chem*, 274(36):25945–25952, Sep 1999.

[174] R. K. Wangsness and F. Bloch. The dynamical theory of nuclear induction. *Phys. Rev.*, 89:728–739, Feb 1953.

[175] D. J. Warren and L. S. Otieno. Carpal tunnel syndrome in patients on intermittent haemodialysis. *Postgrad Med J*, 51(597):450–452, Jul 1975.

[176] Alexander L Watters, Pritilekha Deka, Colin Corrent, David Callender, Gabriele Varani, Tobin Sosnick, and David Baker. The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell*, 128(3):613–624, Feb 2007.

[177] P. Westermark, L. Eriksson, U. Engstr?m, S. Enestr?m, and K. Sletten. Prolactin-derived amyloid in the aging pituitary gland. *Am J Pathol*, 150(1):67–73, Jan 1997.

[178] D. B. Williams, B. H. Barber, R. A. Flavell, and H. Allen. Role of beta 2-microglobulin in the intracellular transport and surface expression of murine class i histocompatibility molecules. *J Immunol*, 142(8):2796–2806, Apr 1989.

[179] Suguru Yamamoto and Fumitake Gejyo. Historical background and clinical treatment of dialysis-related amyloidosis. *Biochim Biophys Acta*, 1753(1):4–10, Nov 2005.